

CC The present sequence represents the N-terminal peptide of a mature
CC outer membrane protein 21 (OMP21) of *Moraxella catarrhalis* strain
CC ATCC49143. The specification describes an OMP21 from a *Moraxella*
CC *catarrhalis* strain, which has an apparent molecular weight of 16-20 kD
CC as determined by sodium dodecylsulfate polyacrylamide gel electrophoresis
CC (SDS-PAGE). OMP21, its nucleic acids and antibodies can be used in
CC prophylactic and therapeutic compositions for treating a *M. catarrhalis*
CC bacterial infection, otitis media, respiratory infections, sinusitis
CC and pneumonia. They are useful as reagents for the clinical or medical
CC diagnosis of *M. catarrhalis* infections and for scientific research on
CC the properties of pathogenicity, virulence and infectivity of
CC *M. catarrhalis* and host defence mechanisms. The antibodies, particularly
CC those that are cytotoxic may be used in passive immunisation to prevent
CC or attenuate *M. catarrhalis* infections of animals e.g. humans.

XX
SQ Sequence 40 AA:

Query Match 100.0%; Score 209; DB 21; Length 40;
Best Local Similarity 100.0%; Pred. No. 8.9e-24;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AISYNSADAQPYVGAKIGVDAKQINGKNTAYGIYAGYN 40
Db 1 aisynsadaqpyvgakigvdakqngkntaygiyagyn 40
|||||

RESULT 2
Y84612
ID Y84612 standard; Protein; 180 AA.
XX
AC Y84612;
XX
DT 25-JUL-2000 (first entry)
XX
DE The outer membrane protein 21 of strain ATCC49143.
XX
KW Outer membrane protein 21; OMP21; strain ATCC49143; bacterial infection;
KM otitis media; respiratory infection; sinusitis; pneumonia; immunisation.
XX
OS *Moraxella catarrhalis*.
XX
FH Key Location/Qualifiers
FT Misc-difference 71 /note- "Ala encoded by CCC"
XX
PN WO200018910-A1.
XX
PD 06-APR-2000.
XX
PE 01-OCT-1999; 99WO-US22918.
XX
PR 01-OCT-1998; 98US-0164714.
XX
PA (ANTE-) ANTEX BIOLOGICS INC.
XX
PI Tucker K, Tillmann UF;
XX
XX WPI: 2000-293149/25.
DR N-PSDB; A12591.
XX
PT Isolated outer membrane protein from a *Moraxella catarrhalis* strain
PT used for diagnosis treatment and prevention of disease caused by *M.*
PT *catarrhalis* e.g. pneumonia, otitis media and respiratory infections -
XX
PS Claim 2; Fig 4; 108pp; English.
XX
XX The present sequence represents an outer membrane protein 21 (OMP21)
CC of *Moraxella catarrhalis* strain ATCC49143. The OMP21 protein has an
CC apparent molecular weight of 16-20 kD as determined by sodium
CC dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE). OMP21,
CC its nucleic acids and antibodies can be used in prophylactic and
CC therapeutic compositions for treating a *M. catarrhalis* bacterial

CC infection, otitis media, respiratory infections, sinusitis and
CC pneumonia. They are useful as reagents for the clinical or medical
CC diagnosis of *M. catarrhalis* infections and for scientific research on
CC the properties of pathogenicity, virulence and infectivity of
CC *M. catarrhalis* and host defence mechanisms. The antibodies, particularly
CC those that are cytotoxic may be used in passive immunisation to prevent
CC or attenuate *M. catarrhalis* infections of animals e.g. humans.

XX
SQ Sequence 180 AA:

Query Match 100.0%; Score 209; DB 21; Length 180;
Best Local Similarity 100.0%; Pred. No. 5.6e-23;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AISYNSADAQPYVGAKIGVDAKQINGKNTAYGIYAGYN 40
Db 23 aisynsadaqpyvgakigvdakqngkntaygiyagyn 62
|||||

RESULT 3
Y56863
ID Y56863 standard; peptide; 31 AA.
XX
AC Y56863;
XX
DT 10-APR-2000 (first entry)
XX
DE B. *catarrhalis* 20 kD antigen N-terminal fragment.
XX
KW Branhamella catarrhalis; *Moraxella catarrhalis*; antigen; vaccine;
KM immunogenic; respiratory tract infection; otitis media.
XX
OS Branhamella catarrhalis.
XX
PN WO9958563-A2.
XX
PD 18-NOV-1999.
XX
PE 11-MAY-1999; 99WO-GB01473.
XX
PR 11-MAY-1998; 98GB-0010084.
XX
PA (CORT-) CORTECS UK LTD.
XX
PI Cripps AW, Kyd J;
XX
XX WPI: 2000-116286/10.
XX
DR
XX
PT Novel antigens of Branhamella catarrhalis used for diagnosis, detection
PT and in vaccines -
XX
PS Claim 5; Page 22; 32pp; English.
XX
XX The invention relates to novel Branhamella catarrhalis (*Moraxella*
CC *catarrhalis*) antigens having an apparent molecular weight of about 14-71
CC kDa (as determined by SDS- PAGE). The antigens can be used to prepare
CC vaccines and immunogenic compositions for the treatment and prophylaxis
CC of Branhamella catarrhalis infections, respiratory tract infections, and
CC otitis media. Antibodies against the antigens can be used for diagnosis
CC and purification of the antigens. Sequences Y56863-871 represent peptide
CC peptide fragments from the B. *catarrhalis* antigens of the invention.
XX
SQ Sequence 31 AA:

Query Match 72.7%; Score 152; DB 21; Length 31;
Best Local Similarity 96.8%; Pred. No. 1.4e-15;
Matches 30; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AISYNSADAQPYVGAKIGVDAKQINGKNT 31
Db 1 aisynsadaqpyvgakigvdakqinknt 31
|||||

```

RESULT 4
ID W79196 Standard; Protein; 680 AA.
XX
AC W79196;
XX
DT 19-JUN-2000 (first entry)
XX
DE Maize glutamine:fructose-6-phosphate amidotransferase.
XX
KM glutamine:fructose-6-phosphate aminotransferase; GFAT; maize;
XX
KW transgenic plant; cationic starch.
XX
OS Zea mays.
XX
PN W0200011192.72.
XX
PD 02-MAR-2000.
XX
PE 19-AUG-1999; 99WO-US18789.
XX
PR 25-AUG-1998; 98US-0097881.
XX
PA (P10N-) PIONEER III-BRED INT INC.
XX
PI Nichols SE, Wang T, Dhugra KS, Wang X, Fallis PL, Bowen BA;
XX
DR WPI: 2000-237656/20.
XX
N-PSDB: Z94235.
XX
PT New plant glutamine:fructose-6-phosphate amidotransferase protein for
PT producing transgenic plants containing cationic starch useful in the
PT textile and paper industries -
XX
PS Claim 24; Page 65-66; 68pp; English.
XX
CC The present sequence is that of maize glutamine:fructose-6-phosphate
CC amidotransferase (GFAT), an enzyme that catalyses the rate-limiting
CC step of the hexosamine biosynthetic pathway, i.e. the formation of
CC glucosamine-6-phosphate and glutamate from fructose-6-phosphate and
CC glutamine. Expression of GFAT in transgenic plants results in
CC production of cationic starch comprising 2-amino anhydroglucose
CC moieties. The invention provides GFAT polypeptides and nucleic
CC acids, as well as vectors capable of expressing the nucleic acids,
CC host cells containing the vectors, and transgenic plants and seeds
CC that produce cationic starch. Suitable plants are selected from
CC maize, sorghum, wheat, rice, barley, oat, sunflower, soybean,
CC saffron, Brassica, cassava, sweet potato and potato. Cationic
CC starch is used in textile sizing, adhesives, water purification,
CC detergents and paper manufacture. The invention precludes the need
CC for chemical modification processes involved in prior art procedures
CC for cationic starch manufacture.
XX
SQ Sequence 680 AA;
XX
Query Match 27.3%; Score 57; DB 21; Length 680;
Best Local Similarity 32.0%; Pred. No. 4.4;
Matches 16; Conservative 9; Mismatches 7; Indels 18; Gaps 3
OY 6 NSADADPVG-----KT-----GQVDAKQJNGKMTAVCIYAG 38
Db 55 ssdardpyagappjlvfrgeqkientlvrsysevdekdvn-ldafsrvhag 103
RESULT 5
W21572
ID W21572 standard; Protein; 1235 AA.
AC W21572;
XX

```

DT	20-AUG-1997	(first entry)
XX		
DE	Rat brain serine ectopeptidase.	
XX		
XX	Serine ectopeptidase; tripeptidylpeptidase II; 'pp II;	
XX	alternative splicing; rodent; rat; cerebral cortex;	
KW	neuropeptide; cholecystokinin; CCK; inactivation; degradation;	
KW	anorexia; schizophrenia; Parkinson's disease; depression;	
KW	irritable bowel syndrome; bulimia; pathological obesity.	
XX		
XX	Rattus sp.	
OS		
XX		
FH	Key	Location/Qualifiers
FT	Region	85..1235
FT		/note="The C-terminal region of the protein encoded
FT		by the alternatively spliced clone has a
FT		sequence identical to this region"
XX		
PN	W09635805-A2.	
XX		
PD	14-NOV-1996.	
XX		
PF	09-MAY-1996;	96WO-FR00700.
XX		
PR	09-MAY-1995;	95FR-0005489.
XX		
PA	(INRM) INSERM INST NAT SANTE & RECH MEDICALE.	
PI	Banbaj RB, Bishop PB, Bourgeat P, Chan S, Ganellin CR;	
PI	Lehland B, Moore AMJ, Schwartz JC, Vargas F;	
PI	Linha Z, Rose C;	
XX		
DR	WPI: 1996-518693/51.	
DR	N-PSDB; T72654.	
XX		
PT	Screening medicaments for treating disorders linked to inactivation	
PT	of endogenous neuro-peptide(s) - by contacting candidate molecule	
PT	with membrane tri:peptidyl-peptidase (homologue) and measuring	
PT	enzyme activity	
XX		
PS	Claim 4; Pages 166-171; 212pp; French.	
XX		
CC	An enzyme with specificity for cholecystokinin (CCK) substrates	
CC	(specifically the non-sulphated CCK8 and the CCK5 peptides) was	
CC	purified from rat cerebral cortex membranes using high performance	
CC	liquid chromatography. Based on amino acid sequence data from the	
CC	purified enzyme, PCR primers were designed to amplify oligonucleotide	
CC	probes A and B of lengths 350 and 380 nucleotides, respectively.	
CC	Using these probes, two distinct clones were identified in a rat	
CC	brain cDNA library. The first clone encoded the protein which is	
CC	the rodent homologue of human tripeptidylpeptidase II. The second	
CC	clone differed in the 5' region, probably as a result of alternative	
CC	splicing, and encodes a protein of the present sequence which appears	
CC	to be a serine ectopeptidase. The rat enzymes are preferred for use	
CC	in a new method of screening for medicaments for treating disorders	
CC	linked to the inactivation or degradation (or being treatable by	
CC	reversing physiological degradation) of endogenous neuropeptides. In	
CC	the method, a candidate molecule is contacted with a membrane	
CC	tripeptidylpeptidase or homologue, and enzyme activity is measured.	
CC	Disorders linked to inactivation or degradation of endogenous	
CC	neuropeptides, include food intake disorders, cognitive and motor	
CC	disorders such as anorexia, schizophrenia, Parkinson's disease and	
CC	depression, as well as gastrointestinal transit disorders such as	
CC	irritable bowel syndrome, bulimia and pathological obesity.	
XX		
XX		
SO	Sequence	1235 AA;

Query Match 25.4%; Score 53; DB 17; Length 1235;

Best Local Similarity 36.7%; Pred No. 35;

Matches 11; Conservative 6; Mismatches 13; Indels 0; Gaps 0;

10 AOPYGAKTGVDAKQINCKNTAYGIYACY 39

Db 1124 aqphdgaagdaakeegestleslsety 1153

```

||||| 1 1 :||| 1:1 1
RESULT 6
W21571
ID W21571 standard; Protein; 1249 AA.
XX
AC W21571:
XX
DT 20-AUG-1997 (first entry)
XX
DE Rat brain homologue of human tripeptidylpeptidase II.
XX
KW Tripeptidylpeptidase II; TRP II; rodent; rat; cerebral cortex;
KW neuropeptide; cholecystokinin; CCK; inactivation; degradation;
KW anorexia; schizophrenia; Parkinson's disease; depression;
KW irritable bowel syndrome; bulimia; pathological obesity;
KW alternative splicing.
XX
OS Rattus sp.
XX
FH Key Location/Qualifiers
FT Region 99..1249
FT /note- "The C-terminal region of the protein encoded
FT by the alternatively spliced clone has a
FT sequence identical to this region"
FT
XX
PN M09635805.A2.
XX
XX 14-NOV-1996.
XX
PD 09-MAY-1996; 96MO-FR00700.
XX
PE 09-MAY-1995; 95FR-0005489.
XX
PR 09-MAY-1995; 95FR-0005489.
XX
PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI Bambaï RB, Bishop PB, Bourgeat P, Chan S, Ganellin CR;
PI Loblon B, Moore ANJ, Schwartz JC, Vargas F;
PI Lihua Z, Rose C;
XX
DR WPI: 1996-518693/51.
DR N-PSDB: T72653.
XX
PT Screening medicaments for treating disorders linked to inactivation
PT of endogenous neuro-peptide(s) by contacting candidate molecule
PT with membrane tri-peptidyl-peptidase (homologue) and measuring
PT enzyme activity
XX
PS Claim 3; Pages 161-165; 212pp; French.
XX
CC An enzyme with specificity for cholecystokinin (CCK) substrates
CC (specifically the non-sulphated CCK8 and the CCK5 peptides) was
CC purified from rat cerebral cortex membranes using high performance
CC liquid chromatography. Based on amino acid sequence data from the
CC purified enzyme, PCR primers were designed to amplify oligonucleotide
CC probes A and B of lengths 350 and 380 nucleotides, respectively.
CC Using these probes, two distinct clones were identified in a rat
CC brain cDNA library. The first clone encoded the protein of the
CC present sequence which is the rodent homologue of human tripeptidyl-
CC peptidase II. The second clone differed in the 5' region, probably
CC as a result of alternative splicing, and appears to be a serine
CC ectopeptidase. The rat enzymes are preferred for use in a new
CC method of screening for medicaments for treating disorders linked
CC to the inactivation or degradation (or being treatable by retarding
CC physiological degradation) of endogenous neuropeptides. In the
CC method, a candidate molecule is contacted with a membrane tripeptidyl-
CC peptidase or homologue, and enzyme activity is measured.
CC Disorders linked to inactivation or degradation of endogenous
CC neuropeptides, include food intake disorders, cognitive and motor
CC disorders such as anorexia, schizophrenia, Parkinson's disease and
CC depression, as well as gastrointestinal transit disorders such as

```

CC irritable bowel syndrome, bulimia and pathological obesity.
 XX
 SQ Sequence 1249 AA;

Query Match 25.4%; Score 53; DB 17; Length 1249;
 Best Local Similarity 36.7%; Pred. No. 36;
 Matches 11; Conservative 6; Mismatches 13; Indels 0; Gaps 0;

QY 10 AOPYVGAKIGQVDAKQINGKNPAGYVAGY 39
 ||| 1 1 :||| 1:1 1
 Db 1138 aqphdgaagdaakeegestleslsety 1167

```

RESULT 7
R34398
ID R34398 standard; Protein; 199 AA.
XX
AC R34398:
XX
DT 03-AUG-1993 (first entry)
XX
DE Helicobacter pylori ureg gene.
XX
KW ureg; uref; ureh; urei; stomach; gastric mucosa; gastric ulcer;
KW duodenal ulcer; acid tolerance; urease operon.
XX
OS Helicobacter pylori.
XX
FH Key Location/Qualifiers
FT Region 8..15
FT /note- "ATP-binding"
FT
XX
PN M09307273-A.
XX
PD 15-APR-1993.
XX
PE 02-OCT-1992; 92MO-FR00921.
XX
PR 03-OCT-1991; 91FR-0012198.
XX
PA (INRM ) INSERM INST NAT SANTE & RECH MED.
PA (INSP ) INST PASTEUR.
XX
PI Cussac V, Ferrero R, Labigne A;
XX
DR WPI: 1993-134462/16.
DR N-PSDB: Q40312.
XX
PT Helicobacter pylori genes useful in diagnosis, vaccines and
PT treatment - necessary for the regulation and maturation of urease
XX
PS Claim 18; Fig 4; 94pp; French.
XX
CC The ureg gene is one of 5 new urease genes to be identified by
CC deletion studies in E.coli where all 5 were found to be necessary
CC for functional urease expression (regulation/maturation). The ureg
CC gene encodes a protein of estimated mol. wt. 21.7kD. The ureg
CC polypeptide of H.pylori has 92% conservation and 59% identity with
CC the ureg polypeptide of K.aerogenes. In K.aerogenes the ureg gene
CC is implicated in the activation of the apoenzyme by incorporation
CC of nickel in the urease subunits. See also R34395-R34399.
XX
SQ Sequence 199 AA;

```

Query Match 24.9%; Score 52; DB 14; Length 199;
 Best Local Similarity 40.0%; Pred. No. 5.3;
 Matches 12; Conservative 6; Mismatches 10; Indels 2; Gaps 1;
 QY 6 NSADAPYVGA--KIGQVDAKQINGKNYAY 33
 | | ||||| 1:1 :||| 1:1 :
 Db 145 nkldlapyvgadlkvmerdskkiaksp1f 174

RESULT 8
 ID Y05734 standard; Protein; 1074 AA.
 AC Y05734;
 DT 19-JUL-1999 (first entry)
 DE Human Grb7 effector 2.2412 protein.
 KM Grb7 effector; 2.2412 protein; human; signal transduction;
 KW tumour marker; breast cancer; prostate cancer; prognosis;
 KW diagnosis.
 XX Homo sapiens.
 OS
 PN M09915647-A1.
 PD 01-APR-1999.
 PF 23-SEP-1998; 98WO-AU00795.
 PR 23-SEP-1997; 97AU-0009388.
 PA (GARV-) GARVAN INST MEDICAL RES.
 PI Daly RJ, Sutherland RL,
 XX
 DR WPI: 1999-254707/21.
 DR N-PSDB: X25366.
 PT New candidate effector for the Grb7 family of signaling proteins,
 PT and specific antibody, useful for detection and treatment of cancer
 XX
 PS Claim 9; Fig 1; 24pp; English.
 XX
 CC The present sequence represents a novel candidate effector for
 CC the Grb7 family of signalling proteins, termed 2.2412. The
 CC sequence is predicted from a partial cDNA (see X25366). 2.2412
 CC mRNA transcripts (7 kb) were detected in all tissues examined
 CC with the exception of kidney. Expression was particularly high in
 CC skeletal muscle and placenta. Analysis of the sequence revealed
 CC significant homology to a large number of proteins containing
 CC ankyrin-like repeats. The 2.2412 gene was localised to between
 CC chromosome 10q23.2 and proximal 10q23.32. Deletions in the
 CC 10q23-25 region have been detected in human breast, prostate,
 CC renal, small cell lung and endometrial carcinomas, glioblastoma
 CC multiforme, melanoma and meningiomas. Detection of the protein
 CC encoded by the 2.2412 cDNA in a sample should provide a useful
 CC tumour marker and/or prognostic indicator for certain human
 CC cancers, in particular breast cancer and prostate cancer.
 CC Antagonism of the interaction between Grb7 family members and the
 CC encoded protein should provide a novel treatment strategy for human
 CC diseases exhibiting aberrant receptor tyrosine kinase signalling,
 CC such as cancer. Anti-Grb7 antibodies can be used in methods of
 CC detecting the presence of 2.2412 protein in a sample.
 XX
 Sequence 1074 AA:

	Query Match	Similarity	24.4%	Score 51:	DB 20;	Length 1074;
	Best Local	Similarity	35.8%	Pred. No. 58;		
Matches	19;	Conservative	4;	Mismatches	16;	Indels 14; Gaps 3.
QY	2	TTCGMS-ADADPGVCAKIGQVD-----AKQINGK-NTRYGIYAGYN	40			
		: : :				
Db	393	islgnseadrgllleaakagdyetvkkllctygsvnncrdlegirstplhfagaayn	445			
RESULT	9					
	A37398					

ID	Y37398 standard; Protein; 221 AA.
XX	
AC	Y37398.
XX	
DT	07-OCT-1999 (first entry)
XX	
DE	Protein involved in intermediate metabolism of sugars and/or cofactors.
XX	
KW	Vaccine; eye disease; conjunctivitis; nonendemic trachoma;
KW	paratrachoma; inclusion conjunctivitis; genital disease; perithenaritis;
KW	nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
KW	bartholinitis; pneumopathy; venereal lymphogranulomatosis.
XX	
CS	Chlamydia trachomatis.
PN	W09928475-A2.
PD	10-JUN-1999.
XX	
PE	27-NOV-1998; 98W0-IB01939.
XX	
PR	04-NOV-1998; 98U0-0107077.
PR	28-NOV-1997; 97FR-0015041.
PR	17-DEC-1997; 97FR-0016034.
XX	
PA	(GEST) GENSET.
XX	
PI	Griffais R;
XX	
DR	WPI; 1999-371125/31.
XX	
PT	Genome sequence of Chlamydia trachomatis
XX	
PS	Disclosure: Page 1105-1106; 1755pp; English.
XX	
CC	Y36754-Y37949 are encoded by open reading frames (ORFs) of the genome of
CC	Chlamydia trachomatis (see Z01425). The polypeptides can be used as
CC	vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences
CC	can also be used to control growth of the microorganism. Chlamydia
CC	trachomatis is responsible for a large number of diseases, e.g. eye
CC	diseases such as conventional trachoma, nonendemic trachoma,
CC	paratrachoma, and inclusion conjunctivitis; genital diseases such as
CC	nongonococcal urethritis, epididymitis, cervicitis, salpingitis,
CC	perithenaritis, bartholinitis, pneumopathy in breast feeding infants;
CC	and venereal lymphogranulomatosis. The polypeptides of the invention
CC	may be of use in treating these diseases.
XX	
Sequence	221 AA;

Query Match	24.2%	Score 50.5	DB 20	Length 221
Best Local Similarity	38.2%	Pred. No. 9.9		
Matches 13	Conservative 3	Mismatches 15	Indels 3	Gaps 1
QY	5 GNSADAPYVG---AKIGOVDAKQINGKNTAYGI 35			
Db	99 ggsadlssdglwlaiekgllssdfllgnirxygv 132			
RESULT 10				
W27148				
ID W27148	standard; Protein; 833 AA.			
XX				
AC W27148;				
XX				
DT 04-DEC-1997	(first entry)			
XX				
DE HMG-CoA reductase degradation protein Hrd3p.				
XX				
KW 3-hydroxy-3-methylglutaryl: coenzyme A; cholesterol; Hrdip; Hrd3p;				
RW Hrd3p; hypercholesterolaemia; yeast.				
XX				
XX Saccharomyces cerevisiae.				

```

XX  Key                               Location/Qualifiers
FH  Peptide                          1..18
FT  /note= "Putative signal sequence"
FT  Region                          768..789
FT  /note= "Strongly hydrophobic"
XX  W09707219-A2.
XX
XX  27-FEB-1997.
XX
XX  16-AUG-1996; 96MO-IB01161.
XX
XX  17-AUG-1995; 95US-0002381.
XX
XX  (REGC ) UNIV CALIFORNIA.
XX
XX  Hampton R, Rine JD;
XX
XX  WPI; 1997-165303/15.
XX
XX  N-PSDB; T85268.
XX
XX  3-hydroxy-3-methylglutaryl CoA reductase degradation polypeptide(s)
XX  - useful as therapeutic agents to reduce hypercholesterolaemia
XX
XX  Claim 4; Page 85-87; 132pp; English.
XX
XX  The present sequence represents the 3-hydroxy-3-methylglutaryl (HMG)-CoA
XX  reductase degradation (HRD) protein Hrdp. Hrd proteins can be used to
XX  regulate the degradation of HMG-CoA reductase, e.g. as therapeutic
XX  agents to reduce hypercholesterolaemia, and to elucidate how the
XX  cholesterol pathway modules the degradation of HMG-CoA reductase. In
XX  addition, as a result of their ability to bind the proteasome complex,
XX  antibodies that specifically bind Hrd polypeptides can be used to
XX  isolate the proteasome complex. Further, they can be used in various
XX  assays to identify compounds that modify the degradation of HMG-CoA
XX  reductase independently of the beneficial LDL receptor control axis.
XX  The nucleic acid molecules can be used as molecular probes for the
XX  isolation of homologous nucleic acid molecules and for the detection of
XX  HRD nucleic acid molecules in yeast.
XX
XX  Sequence 833 AA;
XX
XX  Query Match 23.9%; Score 50; DB 18; Length 833;
XX  Best Local Similarity 37.8%; Pred. No. 60;
XX  Matches 14; Conservative 7; Mismatches 10; Indels 6; Gaps 2;
XX
OY 6 NSADAQPIYV--GAKIGVDAKQINGKNTAYGIVAGYN 40
DB 170 dsakallyygraaglygnikakqv---laykysgfn 202
XX
XX  RESULT 11
XX  W53477
XX  ID W53477 standard; Protein; 255 AA.
XX
XX  W55477;
XX
XX  30-JUN-1998 (first entry)
XX
XX  H. pylori ORF 07ep11916_5273452_c3_31 cell envelope OMP.
XX
XX  Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
XX  identification; binding compound; bacteria; life cycle; activator;
XX  inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;
XX  bacterium.
XX
XX  Helicobacter pylori.
XX
XX  W09737044-A1.
XX
XX  09-OCT-1997.
XX

```

```

XX  27-MAR-1997; 97WO-US05223.
XX
XX  06-DEC-1996; 96US-0761318.
XX  29-MAR-1996; 96US-0625811.
XX  02-APR-1996; 96US-0758731.
XX  25-OCT-1996; 96US-0736905.
XX  28-OCT-1996; 96US-0738859.
XX
XX  (ASTR ) ASTRA AB.
XX
XX  Alm RA, Smith D;
XX
XX  WPI; 1997-503122/46.
XX
XX  N-PSDB; V24886.
XX
XX  Helicobacter pylori nucleic acid sequences and encoded
XX  polypeptide(s) - useful in vaccines to treat or prevent H. pylori
XX  infection and for diagnosis of H. pylori infection
XX
XX  Claims 14,80; Pages 684-685; 1145pp; English.
XX
XX  This sequence is a H. pylori cell envelope outer membrane
XX  protein (OMP) having a C-terminal tyrosine cluster motif.
XX  The protein may be used in a vaccine to prevent or treat H. pylori
XX  infection or to identify H. pylori polypeptide binding compounds,
XX  useful as potential H. pylori life cycle activators or inhibitors.
XX  The DNA and probes derived from it may be used for the
XX  identification of H. pylori in a sample, and the diagnosis of
XX  H. pylori infection. Nucleic acid sequences complementary to the
XX  DNA act as antisense sequences, and can be used to prevent the
XX  translation of H. pylori mRNA. Antibodies against the protein can
XX  be used in immunoassays to evaluate the abundance and distribution
XX  of H. pylori-specific antigens. The genomic sequence of H. pylori
XX  (ATCC 55679) was determined from overlapping contigs generated
XX  by mechanically shearing the bacterial DNA. The sequences were
XX  analysed for ORF of at least 180 nucleotides, and the predicted
XX  coding regions defined by computer evaluation. To identify likely
XX  H. pylori antigens for vaccine development, the amino acid
XX  sequences predicted from various ORF were analysed for significant
XX  homology to other known or exported membrane proteins. Having
XX  identified and determined the sequences of interest, particular
XX  regions can be isolated from H. pylori by PCR amplification for
XX  recombinant polypeptide production, e.g. in E. coli hosts.
XX
XX  Sequence 255 AA;
XX
XX  Query Match 23.7%; Score 49.5; DB 18; Length 255;
XX  Best Local Similarity 34.6%; Pred. No. 16;
XX  Matches 18; Conservative 8; Mismatches 11; Indels 15; Gaps 5;
XX
OY 3 SYG-----NSADAQPIYVAKIGVD-AKQIN-----GKNTAYGIV--AGYN 40
DB 101 tygyysynhahls-fvsgsklgmdgasqvnlfyvgvdalynfyeskegyn 151
XX
XX  RESULT 12
XX  W55201
XX  ID W55201 standard; Protein; 255 AA.
XX
XX  W55201;
XX
XX  15-JUN-1998 (first entry)
XX
XX  H. pylori ORF 01cp11414orf2 protein.
XX
XX  Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
XX  identification; binding compound; bacteria; life cycle; activator;
XX  inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
XX
XX  Helicobacter pylori.
XX

```

NN		M09737044-A1.	
XX			
PD		09-OCT-1997.	
XX			
PF		27-MAR-1997;	97WO-US05223.
XX			
XX		06-DEC-1986;	96US-0761318.
PR		29-MAR-1996;	96US-0625811.
PR		02-APR-1996;	96US-0758731.
PR		25-OCT-1996;	96US-0736905.
PR		28-OCT-1996;	96US-0738859.
XX			
PA		(ASTR) ASTRA AB.	
XI			
PI		Alm RA, Smith D;	
DR			
DR		MP1: 1997-503122/46.	
DR		N-P5DB: V24610.	
XX			
PT		Helicobacter pylori nucleic acid sequences and encoded	
PT		polypeptide(s) - useful in vaccines to treat or prevent H. pylori	
PT		infection and for diagnosis of H. pylori infection	
XX			
PS		Claim 14; Page 453; 1145pp: English.	
CC		This sequence is a H. pylori protein of unspecified function.	
CC		The protein may be used in a vaccine to prevent or treat H. pylori	
CC		infection or to identify H. pylori polypeptide binding compounds,	
CC		useful as potential H. pylori life cycle activators or inhibitors. The	
CC		DNA and probes derived from it may be used for the identification of	
CC		H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic	
CC		acid sequences complementary to the DNA act as antisense sequences and	
CC		can be used to prevent the translation of H. pylori mRNA. Antibodies	
CC		against the protein can be used in immunoassays to evaluate the abundance	
CC		and distribution of H. pylori-specific antigens. The genomic sequence of	
CC		H. pylori (ATCC 55679) was determined from overlapping contigs generated	
CC		by mechanically shearing the bacterial DNA. The sequences were analysed	
CC		for ORF of at least 180 nucleotides, and the predicted coding regions	
CC		defined by computer evaluation. To identify likely H. pylori antigens for	
CC		vaccine development, the amino acid sequences predicted from various ORF	
CC		were analysed for significant homology to other known or exported	
CC		membrane proteins. Having identified and determined the sequences of	
CC		interest, particular regions can be isolated from H. pylori by PCR	
CC		amplification for recombinant polypeptide production, e.g. in E. coli	
CC		hosts.	
XX			
SQ		Sequence	255 AA;
Query Match		23.7%;	Score 49.5; DB 18; Length 255;
Best Local Similarity		34.6%;	Pred. No. 16;
Matches	18;	Conservative	8; Mismatches 11; Indels 15; Gaps 5
OY		3 SYC-----NSADAQPYVGAKIGVD-AKQIN-----GKNFAVGIY--AGYN 40	
	:	: :	
Db		101 tygysynhanis-fvgskljmdgasymnftvgfdalynfyfeskegn 151	
RESULT	13		
ID	Y17205		
XX		Y17205 standard; Protein: 255 AA.	
AC	Y17205:		
DT	03-AUG-1999	(first entry)	
XX			
DE	H. pylori outer membrane polypeptide.		
KM	Outer membrane polypeptide; OMP; vaccine: H. pylori infection; humoral;		
XX	cellular immune response.		
OS	Helicobacter pylori.		

PV	Novel nucleic acid sequence encoding a Candida albicans surface antigen
PT	
XX	
XX	
PD	06-MAY-1999.
XX	
PF	28-OCT-1998; 98WO-US22883.
XX	
PR	17-DEC-1997; 97US-0993001.
XX	
PA	(GENO-) GENOME THERAPEUTICS CORP.
XX	
PI	Alm RA, Ellis RM, Guild BC, Noonan BM, Smith D;
XX	
DR	WPI: 1999-326698/27.
N-PSDB:	X75824.
XX	
PT	Cellular vaccine against Helicobacter pylori
XX	
ES	Claim 7; Page 287-288; 352pp; English.
XX	
CC	The invention relates to a vaccine for preventing or treating infections by Helicobacter pylori. The vaccine contains at least one isolated H. pylori polypeptide, or its fragments, in a carrier, where the carrier is a Salmonella, Vibrio cholerae or Shigella vector containing a nucleic acid encoding the H. pylori polypeptide. The vaccines induce humoral and cellular immune responses. The vaccines are used to treat or prevent infections by H. pylori. Sequences X75779 to X75837 represent nucleic acid sequences encoding H. pylori outer membrane polypeptides (OMPs) Y11760 to Y117218.
CC	
CC	
SO	Sequence 255 AA:
Query Match	23.7%; Score 49.5; DB 20; Length 255;
Best Local Similarity	34.6%; Pred No. 16;
Matches 18;	Conservative 8; Mismatches 11; Indels 15; Gaps 5
OY	3 SYG----NSADAQPYVGAKIGVD-AKQIN-----GKNFAYGIY---AGYN 40 : : : Db 101 tyggyssynmanis-fvskslglmdgsaynftfygvfdalynfyeskegyn 151
RESULT 14	
YB3275	
ID	YB3275 standard; Protein; 1203 AA.
XX	
AC	YB3275;
XX	
DT	16-AUG-2000 (first entry)
XX	
DE	Candida albicans CSA1 surface protein.
XX	
KM	Yeast; infection; detection; treatment; vaccine; antigen; CSA1; probe; primer; vaginitis; immunoglobulin G; IgG.
OS	Candida albicans.
PN	CA2274984-A1.
XX	
PD	10-JAN-2000.
XX	
PE	09-JUL-1999; 99CA-2274984.
XX	
PR	10-JUL-1998; 98CA-2237134.
XX	
PA	(UYLA-) UNTIV LAVAL.
XX	
PI	Bourbonnais Y, Deslauniers N;
XX	
DR	WPI: 2000-318497/28.
N-PSDB:	Z93782.

PT useful for detection, prophylaxis and treatment of candidiasis
 XX Claim 5; Page 32-36; 50pp; English.
 CC The CSA1 surface antigen is unique to *Candida albicans* and so can
 CC be exploited for the treatment of *Candida albicans* infections.
 CC Nucleic acids, probes and primers specific to the CSA1 coding
 CC sequence can be used to detect *C. albicans* infections in patients
 CC by binding to *C. albicans* CSA1 DNA or mRNA in a patient sample. The
 CC CSA1 protein, vaccines comprising it or antibodies directed against
 CC it can be used to treat *C. albicans* infections. The virulence of
 CC infection by *C. albicans* is reduced by binding of the antibody to
 CC the surface antigen and masking it. Recurrent vaginitis in
 CC immunocompetent patients can be created by repeated administration
 CC of CSA1 antigen to stimulate the production of immunoglobulin G
 CC antibodies. The antibodies raised against CSA1 are specific for *C.*
 CC *albicans* CSA1 antigen but not to other yeast or mycelial antigens,
 CC making detection and treatment easier.
 XX
 SQ Sequence 1203 AA;

Query Match 23.7%; Score 49.5; DB 21; Length 1203;
 Best Local Similarity 48.3%; Pred. No. 1.1e+02;
 Matches 14; Conservative 0; Mismatches 12; Indels 3; Gaps 1;

OY 3 SYGNSADAOPYVGAKIGVDKQINGKNT 31
 ||||| ||| | | | | |
 Db 951 sygnstiaqpstsk---sdaasttgpit 976

RESULT 15
 Y84814
 Y84814 standard; protein; 1203 AA.

AC Y84814;
 XX
 DT 08-AUG-2000 (first entry)

DE Amino acid sequence of the CSA1 gene of *Candida albicans*.

KW CSA1 gene; surface antigen; yeast; passive immunization;
 KW *Candida albicans* infection; vaginal infection; systemic infection;
 KW anti-fungal.

XX
 OS *Candida albicans*.

XX Key Location/Qualifiers
 FH Peptide 1..17
 FT /note= "signal peptide"
 FT Domain 42..143
 FT /note= "C-rich domain"
 FT Domain 233..334
 FT /note= "C-rich domain"
 FT Domain 403..504
 FT /note= "C-rich domain"
 FT Domain 565..666
 FT /note= "C-rich domain"
 FT Domain 750..851
 FT /note= "C-rich domain"
 FT Modified-site 883
 FT /note= "putative N-glycosylation site"
 FT Modified-site 893
 FT /note= "putative N-glycosylation site"
 FT Modified-site 903
 FT /note= "putative N-glycosylation site"
 FT Modified-site 913
 FT /note= "putative N-glycosylation site"
 FT Modified-site 914
 FT /note= "putative N-glycosylation site"
 FT Modified-site 927
 FT /note= "putative N-glycosylation site"
 FT Modified-site 928

FT Modified-site /note= "putative N-glycosylation site"
 FT 937
 FT /note= "putative N-glycosylation site"
 FT Modified-site 938
 FT /note= "putative N-glycosylation site"
 FT Modified-site 954
 FT /note= "putative N-glycosylation site"
 FT Peptide 1184..1203
 FT /note= "predicted GPI-anchoring determinant"

CA2237134-A1.

10-JAN-2000.

10-JUL-1998; 98CA-2237134.

10-JUL-1998; 98CA-2237134.

(UYLA-) UNIV LAVAL.

Deslauriers N, Bourbonnais Y;

WPI; 2000-293503/26.

PT New CSA1 gene coding for a *Candida albicans* surface antigen, useful for
 PT diagnosing the yeast form of *Candida albicans* within biological samples
 PT by immunoassay or polymerase chain reaction -
 PS Disclosure; Fig 4; 36pp; English.

XX The present sequence represents a protein encoded by the CSA1 gene.
 CC The protein is a *Candida albicans* surface antigen. The CSA1 gene is
 CC an intronless single copy gene in *Candida*, and RNA transcripts can
 CC be detected in exponentially growing yeast cells. CSA1 gene expression
 CC is strongly increased upon induction of the mycelial growth phase.
 CC The CSA1 gene, its protein and antibodies against it are useful for
 CC diagnosing the yeast form of *Candida albicans* within biological samples
 CC by immunoassay or polymerase chain reaction (PCR). The antibody is
 CC useful for passive immunization against *Candida albicans* infection.
 CC The antibody is also useful in the treatment of vaginal or systemic
 CC infection by *Candida*. The protein is useful as a therapeutic target
 CC in the development of anti-fungal agents.

SQ Sequence 1203 AA;

Query Match 23.7%; Score 49.5; DB 21; Length 1203;
 Best Local Similarity 48.3%; Pred. No. 1.1e+02;
 Matches 14; Conservative 0; Mismatches 12; Indels 3; Gaps 1;

OY 3 SYGNSADAOPYVGAKIGVDKQINGKNT 31
 ||||| ||| | | | | |
 Db 951 sygnstiaqpstsk---sdaasttgpit 976

Search completed: March 6, 2001, 14:36:12
 Job time: 120 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using SW model

Run on: March 6, 2001, 14:34:37 ; Search time 19.25 Seconds
(without alignments)
37.313 Million cell updates/sec

Title: US-09-164-714-1
Perfect score: 209
Sequence: 1 AISYGNADAPYVGAIGGVDAKQINGKNVAYGIYAGYN 40

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues

Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 10%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2.6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2.6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2.6/ptodata/2/1aa/6.COMB.pep:*
4: /cgn2.6/ptodata/2/1aa/PCUS.COMB.pep:*
5: /cgn2.6/ptodata/2/1aa/backfiles1.pep:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	52	24.9	199	US-08-211-312-6	Sequence 6, Appl1
2	52	24.9	199	US-08-472-285-6	Sequence 6, Appl1
3	50	23.9	833	US-08-699-103B-4	Sequence 4, Appl1
4	48	23.0	532	US-08-911-321-11	Sequence 11, Appl1
5	48	23.0	532	US-08-911-321-11	Sequence 11, Appl1
6	47.5	22.7	283	US-08-852-401-4	Sequence 4, Appl1
7	46	22.0	832	US-08-431-560-1	Sequence 1, Appl1
8	46	22.0	832	US-08-463-345-1	Sequence 1, Appl1
9	46	22.0	1287	US-08-200-232-2	Sequence 1, Appl1
10	46	22.0	1287	US-08-200-232-2	Sequence 1, Appl1
11	46	22.0	1287	US-08-911-321-11	Sequence 11, Appl1
12	46	22.0	1296	US-08-470-260-3	Sequence 3, Appl1
13	46	22.0	1296	US-08-471-491-3	Sequence 3, Appl1
14	46	22.0	1296	US-08-466-662-3	Sequence 3, Appl1
15	45.5	21.8	251	US-08-209-747-8	Sequence 8, Appl1
16	45.5	21.8	251	US-08-458-298-8	Sequence 8, Appl1
17	45.5	21.8	331	US-08-356-180-3	Sequence 3, Appl1
18	45.5	21.8	592	US-08-961-083-222	Sequence 222, App
19	45.5	21.8	2233	US-08-569-853-1	Sequence 1, Appl1
20	45.5	21.8	2233	US-08-569-853-1	Sequence 1, Appl1
21	45.5	21.8	2233	US-08-987-439-1	Sequence 1, Appl1
22	45	21.5	462	US-08-865-597A-2	Sequence 2, Appl1
23	45	21.5	626	US-08-956-242-2	Sequence 2, Appl1
24	45	21.5	626	US-08-956-242-2	Sequence 2, Appl1
25	45	21.5	968	US-08-351-215-2	Sequence 2, Appl1
26	44.5	21.3	132	US-08-961-083-194	Sequence 194, App
27	44.5	21.3	359	US-08-457-997B-2	Sequence 2, Appl1
28	44.5	21.3	359	US-08-467-722A-2	Sequence 2, Appl1

29	44	21.1	20	US-08-462-221-4	Sequence 4, Appl1
30	44	21.1	32	US-08-468-692-21	Sequence 21, Appl1
31	44	21.1	32	US-08-488-351A-21	Sequence 21, Appl1
32	44	21.1	74	US-08-343-443B-11	Sequence 11, Appl1
33	44	21.1	198	US-08-943-915-33	Sequence 33, Appl1
34	44	21.1	207	US-08-943-915-33	Sequence 33, Appl1
35	44	21.1	207	US-08-943-915-33	Sequence 33, Appl1
36	44	21.1	263	US-08-752-844-66	Sequence 66, Appl1
37	44	21.1	339	US-08-266-451B-22	Sequence 22, Appl1
38	44	21.1	339	US-08-748-725-22	Sequence 22, Appl1
39	44	21.1	340	US-08-355-844-1	Sequence 1, Appl1
40	44	21.1	340	US-08-355-844-1	Sequence 1, Appl1
41	44	21.1	494	US-09-198-956-2	Sequence 2, Appl1
42	43.5	20.8	380	US-08-472-659-34	Sequence 34, Appl1
43	43.5	20.8	380	US-08-472-659-34	Sequence 34, Appl1
44	43.5	20.8	380	US-08-474-661-34	Sequence 34, Appl1
45	43.5	20.8	449	US-08-657-392-2	Sequence 2, Appl1

ALIGNMENTS

RESULT 1
US-08-211-312-6
Sequence 6, Application US/08211312
Patent No. 5986051
GENERAL INFORMATION:
APPLICANT: LABIGNE, AGNES
APPLICANT: CUSACK, VALERIE
APPLICANT: FERRERO, RICHARD
TITLE OF INVENTION: GENES OF HELICOBACTER PYLORI NECESSARY
FOR THE REGULATION AND MATURATION OF UREASE AND THEIR USE
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
ADDRESS: P.C.
STREET: 1755 S. Jefferson Davis Highway, Suite 400
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/211,312
FILING DATE: 01-JUL-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 91 12198
FILING DATE: 03-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR92/00921
FILING DATE: 02-OCT-1992
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NO. 5986051man F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 660-075-0XPCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 OPT UR
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 199 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-211-312-6

Query Match 24.9%; Score 52; DB 2; Length 199;
Best Local Similarity 40.0%; Pred. No. 2.4;
Matches 12; Conservative 6; Mismatches 10; Indels 2; Gaps 1;
QY 6 NSADAPYVGA--KIGQVDAKQINKNTAY 33
DB 145 NKIDAPYVGADLKVMERDSKRIAKSPLF 174

RESULT 2
US-08-472-285-6
Sequence 6, Application US/08472285
Patent No. 6027878
GENERAL INFORMATION:
APPLICANT: LABIGNE, AGNES
APPLICANT: CUSSAC, VALERIE
APPLICANT: FERRERO, RICHARD
TITLE OF INVENTION: GENES OF HELICOBACTER PYLORI NECESSARY
TITLE OF INVENTION: FOR THE REGULATION AND MATURATION OF UREASE AND THEIR USE
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
ADDRESS: P.C.
STREET: 1755 S. Jefferson Davis Highway, Suite 400
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/472.285
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/211,312
FILING DATE: 01-JUL-1994
APPLICATION NUMBER: FR 91 12198
FILING DATE: 03-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR92/00921
FILING DATE: 02-OCT-1992
ATTORNEY/AGENT INFORMATION:
NAME: Oblon, No. 6027878man F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 660-075-0XPCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 199 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-472-285-6

Query Match 24.9%; Score 52; DB 3; Length 199;
Best Local Similarity 40.0%; Pred. No. 2.4;
Matches 12; Conservative 6; Mismatches 10; Indels 2; Gaps 1;
QY 6 NSADAPYVGA--KIGQVDAKQINKNTAY 33
DB 145 NKIDAPYVGADLKVMERDSKRIAKSPLF 174

RESULT 3
US-08-699-103B-4

Sequence 4, Application US/08699103B
Patent No. 6107462
GENERAL INFORMATION:
APPLICANT: Rine, Jasper D.
APPLICANT: Hampton, Randolph
TITLE OF INVENTION: GENES AND PROTEINS CONTROLLING
TITLE OF INVENTION: CHOLESTEROL SYNTHESIS
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 2200 Sand Hill Road, Suite 100
CITY: Menlo Park
STATE: CA
COUNTRY: USA
ZIP: 94025
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/699,103B
FILING DATE: 16-AUG-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/002,581
FILING DATE: 17-AUG-1995
ATTORNEY/AGENT INFORMATION:
NAME: Green, Grant D.
REGISTRATION NUMBER: 31,259
REFERENCE/DOCKET NUMBER: 09272/005001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/322-5070
TELEFAX: 650/854-0875
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 833 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-699-103B-4

Query Match 23.9%; Score 50; DB 3; Length 833;
Best Local Similarity 37.8%; Pred. No. 29;
Matches 14; Conservative 7; Mismatches 10; Indels 6; Gaps 2;
QY 6 NSADAPYV--GAKIGQVDAKQINKNTAYGTYAGYN 40
DB 170 DSARKALLYQRAQQLGNKAKQV----LAYRYSGSN 202

RESULT 4
US-08-911-321-11
Sequence 11, Application US/08911321
Patent No. 6010703
GENERAL INFORMATION:
APPLICANT: Roger K. Maes and Stephen J. Spatz
TITLE OF INVENTION: Recombinant Poxvirus
TITLE OF INVENTION: Vaccine Against
TITLE OF INVENTION: Peline Rhinotracheitis
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ian C. McLeod
STREET: 2190 Commons Parkway
CITY: Okemos
STATE: Michigan
COUNTRY: USA
ZIP: 48864
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 KB
MEDIUM TYPE: storage
COMPUTER: IBM compatible
OPERATING SYSTEM: MS-DOS

```
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/911.321
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/096.183
FILING DATE: July 26, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Ian C. McLeod
REGISTRATION NUMBER: 20,931
REFERENCE/DOCKET NUMBER: MSU 4.1-166
TELECOMMUNICATION INFORMATION:
TELEPHONE: (517) 347-4100
TELEFAX: (517) 347-4103
TELEX: NO. 6010703e
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 532
TYPE: Amino Acid
STRANDEDNESS: Single
TOPOLOGY: linear
MOLECULE TYPE:
DESCRIPTION: Polypeptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Feline herpesvirus-1
STRAIN: 1
INDIVIDUAL ISOLATE: C-27
CELL TYPE: N/A
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD: Deduced Sequence
OTHER INFORMATION: 9E
US-08-911-321-11

Query Match          23.0% Score 48; DB 3; Length 532:
Best Local Similarity 36.7%; Pred No. 33;
Matches 11; Conservative 3; Mismatches 16; Indels 0; Gaps 0;

Qy      7 SADAOPTYGAKIGQVDAQINGKNTAYGIY 36
      1 : 11: 11 11 11 11 11:1
Db      299 SINHMPYIEQPNNVDLKFINVPTNASGLY 328

RESULT 5
PCT-US95-13975-2
Sequence 2, Application PC/TUS9513975
GENERAL INFORMATION:
APPLICANT: Cochran, Mark D.
APPLICANT: McDonnell, Michael W.
TITLE OF INVENTION: Recombinant Feline Herpes Virus
NUMBER OF SEQUENCES: 73
CORRESPONDENCE ADDRESS:
ADDRESSEE: John P. White
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM 330 466 DX2
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/13975
FILING DATE: 26-OCT-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
```

```
APPLICATION NUMBER: 08/329,883
FILING DATE: 26-OCT-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 39118-PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)278-0400
TELEFAX: (212)391-0525
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 532 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-13975-2
```

```
Query Match          23.0% Score 48; DB 4; Length 532:
Best Local Similarity 36.7%; Pred No. 33;
Matches 11; Conservative 3; Mismatches 16; Indels 0; Gaps 0;
```

```
Qy      7 SADAOPTYGAKIGQVDAQINGKNTAYGIY 36
      1 : 11: 11 11 11 11 11:1
Db      299 SINHMPYIEQPNNVDLKFINVPTNASGLY 328
```

```
RESULT 6
US-08-852-401-4
Sequence 4, Application US/08852401
Patent No. 5976836
GENERAL INFORMATION:
APPLICANT: Weber, J. Mark
APPLICANT: Hessler, Paul E.
APPLICANT: Larsen, Peter E.
APPLICANT: Luu, B. Minh
TITLE OF INVENTION: Methods and Compositions for Enhancing
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Rockey, Milnamow & Katz, Ltd.
STREET: 2 Prudential Plaza, 180 N. Stetson, Suite
CITY: Chicago
STATE: Illinois
COUNTRY: U.S.A.
ZIP: 60601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/852,401
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mueller, Lisa L.
REGISTRATION NUMBER: 38,978
REFERENCE/DOCKET NUMBER: FER2159POO30US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-616-5400
TELEFAX: 312-616-5460
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 283 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-852-401-4
```



```

: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (510) 601-2708
: TELEFAX: (510) 655-3542
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 1296 amino acids
:   TYPE: amino acid
:   STRANDEDNESS: single
:   TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-470-260-3

Query Match      22.0%; Score 46; DB 3; Length 1296;
Best Local Similarity 40.5%; Pred. No. 2e+02;
Matches 15; Conservative 5; Mismatches 11; Indels 6; Gaps 3;

OY      4 YGNSADAQPYVGAKIGVDAKQINGKNTAVGIYAGYN 40
Db      1049 YGTSAGVDAYLN---GQVEA--IVGGFGSYG-YSSFN 1079

RESULT 13
US-08-471-491-3
: Sequence 3, Application US/08471491B
: Patent No. 6090611
: GENERAL INFORMATION:
: APPLICANT: Covacci, Antonello
: APPLICANT: Bugnoli, Massimo
: APPLICANT: Telford, John
: APPLICANT: Macchia, Giovanni
: APPLICANT: Rappuoli, Rino
: TITLE OF INVENTION: Helicobacter pylori Proteins Useful For Vaccines And
: FILE REFERENCE: CHIR0044
: CURRENT APPLICATION NUMBER: US/08/471.491B
: CURRENT FILING DATE: 1995-06-06
: NUMBER OF SEQ ID NOS: 8
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 3
: LENGTH: 1296
: TYPE: PRT
: ORGANISM: Helicobacter pylori
: US-08-471-491-3

Query Match      22.0%; Score 46; DB 3; Length 1296;
Best Local Similarity 40.5%; Pred. No. 2e+02;
Matches 15; Conservative 5; Mismatches 11; Indels 6; Gaps 3;

OY      4 YGNSADAQPYVGAKIGVDAKQINGKNTAVGIYAGYN 40
Db      1049 YGTSAGVDAYLN---GQVEA--IVGGFGSYG-YSSFN 1079

RESULT 14
US-08-466-662-3
: Sequence 3, Application US/08466662B
: Patent No. 6130059
: GENERAL INFORMATION:
: APPLICANT: Covacci, Antonello
: APPLICANT: Bugnoli, Massimo
: APPLICANT: Telford, John
: APPLICANT: Macchia, Giovanni
: APPLICANT: Rappuoli, Rino
: TITLE OF INVENTION: Helicobacter pylori Proteins Useful For Vaccines And
: FILE REFERENCE: CHIR0057
: CURRENT APPLICATION NUMBER: US/08/466.662B
: CURRENT FILING DATE: 1995-06-06
: NUMBER OF SEQ ID NOS: 8
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 3
```

```

: LENGTH: 1296
: TYPE: PRT
: ORGANISM: Helicobacter pylori
: US-08-466-662-3

Query Match      22.0%; Score 46; DB 3; Length 1296;
Best Local Similarity 40.5%; Pred. No. 2e+02;
Matches 15; Conservative 5; Mismatches 11; Indels 6; Gaps 3;

OY      4 YGNSADAQPYVGAKIGVDAKQINGKNTAVGIYAGYN 40
Db      1049 YGTSAGVDAYLN---GQVEA--IVGGFGSYG-YSSFN 1079

RESULT 15
US-08-209-747-8
: Sequence 8, Application US/08209747
: Patent No. 5733771
: GENERAL INFORMATION:
: APPLICANT: Lewis, Randolph V.
: APPLICANT: Colgin, Mark
: TITLE OF INVENTION: CDNA's Encoding Minor Ampullate Spider
: NUMBER OF SEQUENCES: 56
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Birch, Stewart, Kolasch & Birch
: STREET: P.O. Box 747
: CITY: Falls Church
: STATE: Virginia
: COUNTRY: USA
: ZIP: 22040-3487
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: APPLICATION DATA:
: CURRENT APPLICATION NUMBER: US/08/209.747
: FILING DATE: 14-MAR-1994
: CLASSIFICATION: 530
: ATTORNEY/AGENT INFORMATION:
: NAME: Murphy Jr., Gerald M.
: REGISTRATION NUMBER: 28,977
: REFERENCE/DOCKET NUMBER: 1447-104P
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 703-205-8000
: TELEFAX: 703-205-8050
: INFORMATION FOR SEQ ID NO: 8:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 251 amino acids
:   TYPE: amino acid
:   TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-209-747-8

Query Match      21.8%; Score 45.5; DB 1; Length 251;
Best Local Similarity 35.1%; Pred. No. 30;
Matches 13; Conservative 3; Mismatches 12; Indels 9; Gaps 2;

OY      4 YGNSADAQPYVGAKIGVDAKQINGKNTAVGIYAGYN 40
Db      90 YGCGAGS---GAR-----AASAGAGTGYGSSGCGYN 117

Search completed: March 6, 2001, 14:37:04
Job time: 147 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 14:35:47 ; Search time 23.41 Seconds

(without alignments)
116.020 Million cell updates/sec

Title: US-09-164-714-1

Perfect score: 209
Sequence: 1 A1SYGNSADAQPYVGAKIGVDKQJNGKNTAYGIYAGYN 40

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:.*
2: p1r2:.*
3: p1r3:.*
4: p1r4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	59.5	28.5	380	2 T24786	hypothetical prote
2	59	28.2	160	2 H82207	hypothetical prote
3	55.5	26.6	289	2 S38806	potin opma - Rhodo
4	55.5	26.6	1196	2 S40908	TyB protein - yeas
5	55.5	26.6	1278	2 B22671	TyB protein - yeas
6	55.5	26.6	1328	2 S52481	TyB protein - yeas
7	55.5	26.6	1328	2 S52894	TyB protein - yeas
8	55.5	26.6	1328	2 S50948	TyB protein - yeas
9	55.5	26.6	1328	2 B28097	TyB protein - yeas
10	55.5	26.6	1348	2 B23496	TyB protein - yeas
11	55.5	26.6	1749	2 S69972	TyB protein - yeas
12	55.5	26.6	1755	2 S50641	TyB protein YER138
13	55.5	26.6	1755	2 S50663	TyB protein - yeas
14	55.5	26.6	1755	2 S57047	TyB protein - yeas
15	55.5	26.6	1755	2 S57045	TyB protein - yeas
16	55.5	26.6	1755	2 S45736	TyB protein - yeas
17	55.5	26.6	1755	2 S69963	TyB protein - yeas
18	55.5	26.6	1755	2 S69839	TyB protein - yeas
19	55.5	26.6	1755	2 S69980	TyB protein - yeas
20	55.5	26.6	1755	2 S69957	TyB protein - yeas
21	55.5	26.6	1755	2 S69951	TyB protein - yeas
22	55.5	26.6	1755	2 S69969	TyB protein - yeas
23	55.5	26.6	1755	2 S69979	TyB protein - yeas
24	55.5	26.6	1755	2 S69979	TyB protein - yeas
25	55.5	26.6	1755	2 S69838	TyB protein - yeas
26	55.5	26.6	1755	2 S70298	TyB protein - yeas
27	55.5	26.6	1755	2 S69955	TyB protein - yeas
28	55.5	26.6	1755	2 S69845	TyB protein - yeas
29	55.5	26.6	1755	2 S69866	TyB protein - yeas

30	55.5	26.6	1755	2 S69982	TyB protein - yeas
31	55.5	26.6	1755	2 S69949	TyB protein - yeas
32	55.5	26.6	1755	2 S61763	TyB protein - yeas
33	55.5	26.6	1755	2 S69975	TyB protein - yeas
34	55.5	26.6	1756	2 S45867	TyB protein - yeas
35	55.5	26.6	1756	2 S69983	TyB protein - yeas
36	55.5	26.6	1770	2 S45842	TyB protein - yeas
37	55.5	26.6	1770	2 S69953	TyB protein - yeas
38	55.5	26.6	1770	2 S69948	TyB protein - yeas
39	55.5	26.6	1770	2 S70233	TyB protein - yeas
40	55.5	26.6	1770	2 S70230	TyB protein - yeas
41	55.5	26.6	1770	2 S69966	TyB protein - yeas
42	55.5	26.6	1770	2 S69950	TyB protein - yeas
43	55.5	26.6	1770	2 S58651	TyB protein - yeas
44	55.5	26.6	1771	2 S53592	TyB protein - yeas
45	55.5	26.6	1793	2 S52601	TyB protein - yeas

ALIGNMENTS

```
RESULT 1
T24786
Hypothetical protein ZK1321.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T24786; T27760
R:Gardner, A.
Submitted to the EMBL Data Library, March 1995
A:Reference number: Z19935
A:Accession: T24786
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-380 <W12>
A:Cross-references: EMBL:Z48717; PIDN:CA88612.1; GSPDB:GN00020; CESP:ZK1321.4
A:Experimental source: clone T10B9
R:Gardner, A.
Submitted to the EMBL Data Library, March 1995
A:Reference number: Z20415
A:Accession: T27760
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-380 <W12>
A:Cross-references: EMBL:Z48584; PIDN:CA88478.1; GSPDB:GN00020; CESP:ZK1321.4
A:Experimental source: clone ZK1321
C:Genetics:
A:Gene: CESP:ZK1321.4
A:Map position: 2
A:Introns: 4/3; 35/2; 85/3; 115/2; 224/2; 328/2; 346/3
C:Superfamily: Caenorhabditis elegans hypothetical protein ZK1321.4

Query Match      28.5%; Score 59.5; DB 2; Length 380;
Best Local Similarity 38.6%; Pred. No. 3.1;
Matches 17; Conservative 2; Mismatches 16; Indels 9; Gaps 2;

OY 3 SYGNSADAQPYVGAKIGVDKQJNGKNTAYGIYAG 38
Db 224 STNGATISOPY-GAGSGTPLNOMFTINTSPAPGANGAAGC 266

RESULT 2
H82207
Hypothetical protein VC1384 [imported] - Vibrio cholerae (group O1 strain N16961)
C:Species: Vibrio cholerae
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 15-Sep-2000
C:Accession: H82207
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.;
  chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bess, S.; Qin, H.; Dragol, I.; Sellers
  L.R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: AB2035; MUID:20406833
```

```

A:Accession: J012207
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-160 <HEX>
A:Cross-references: GB:AE004217; GB:AE003852; NID:9655866; PIDN:AAE94542.1; GSPDB:GN001
A:Experimental source: serogroup O1; strain N16961; biotype EI Tor
C:Genetics:
A:Gene: VC1384
A:Map position: 1

Query Match          28.2%; Score 59; DB 2; Length 160;
Best Local Similarity 45.2%; Pred. No. 1.5;
Matches 14; Conservative 6; Mismatches 9; Indels 2; Gaps 1;

OY      8 ADAQPYGAKIGQVDAKOINGKNTAYGIYAG 38
      ||| ||| ||| ||| ||| ||| ||| |||
      19 ADSWITYGASVQSDYEKGKG--TAVSVHAG 47

RESULT      3
S38806
porin opma - Rhodopseudomonas blaslica
C:Species: Rhodopseudomonas blaslica
C:Date: 31-Dec-1993 #sequence_revision 02-Jun-1994 #text_change 07-May-1999
A:Accession: S38806; S38793
R:Kreusch, A.; Neubueser, A.; Schiltz, E.; Weckesser, J.; Schulz, G.E.
Protein Sci. 3, 58-63, 1994
A:Title: Structure of the membrane channel porin from Rhodopseudomonas blaslica at 2.0 A
A:Reference number: S38806; MUID:94191532
A:Accession: S38806
A:Status: nucleic acid sequence not shown; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-289 <KR>
R:Kreusch, A.; Neubueser, A.; Schiltz, E.; Weckesser, J.; Schulz, G.E.
submitted to the Protein Sequence Database, November 1993
A:Reference number: S38793
A:Accession: S38793
A:Molecule type: protein
A:Residues: 1-71, 'X', 73-82; 91-104; 109-137; 146-163; 173-289 <KR>
C:Genetics:
A:Gene: opma
C:Keywords: homotrimer; membrane protein

Query Match          26.6%; Score 55.5; DB 2; Length 289;
Best Local Similarity 41.7%; Pred. No. 8;
Matches 15; Conservative 3; Mismatches 15; Indels 3; Gaps 1;

OY      4 YGNSADQPYGAKIGQVDAKOINGKNTAYGIYAG 39
      ||| ||| ||| ||| ||| ||| ||| |||
      DB 227 YGNVAFGATTVRAYVSDIDRA---GADTAYGIGADY 259

RESULT      4
S40908
TYB protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty1.A
N:Alternate names: protein YAR09C
C:Species: Saccharomyces cerevisiae
C:Date: 13-Jan-1995 #sequence_revision 27-Jan-1995 #text_change 22-Oct-1999
A:Accession: S40908
R:Clark, M.W.; Keng, T.; Storms, R.K.; Zhong, W.; Fortin, N.; Zeng, B.; Delaney, S.; Quecena, S.; Quecena, S.;
Submitted to the EMBL Data Library, November 1993
A:Description: Sequencing of chromosome I of Saccharomyces cerevisiae: Analysis of the 4
A:Reference number: S40891
A:Accession: S40908
A:Molecule type: DNA
A:Residues: 1-1196 <CLAA>
A:Cross-references: EMBL:L22015; NID:g1339990; PIDN:AAC04967.1; PID:g2564964
C:Genetics:
A:Map position: 1R
A:Mobile element: retrotransposon Ty1.A
C:Superfamily: Tyb protein

```

```

Query Match Similarity 26.6%; Score 55.5; DB 2; Length 1196;
Best Local Similarity 43.8%; Pred. No. 33;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

QY 3 SYGNSADAPYVGAKIGOV---DAKQINGKNT 31
      |||| |::|| : | | | |
Db 1053 SYGN---QPYKKSQIGNIYLNGKVIIGKST 1080

RESULT 5
B22671
Tyb protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty912
C:Species: Saccharomyces cerevisiae
C:Date: 23-Aug-1987 #sequence_revision 23-Aug-1987 #text_change 26-Aug-1999
A:Accession: B22671; S05890
R:Clare, J.; Farabaugh, P.
Proc. Natl. Acad. Sci. U.S.A. 82, 2829-2833, 1985
A:Title: Nucleotide sequence of a yeast Ty element: evidence for an unusual mechanism
A:Reference number: A94039; MUID:85190589
A:Accession: B22671
A:Molecule type: DNA
A:Residues: 1-1278 <CUI>
R:McLellor, J.; Fulton, S.M.; Dobson, M.J.; Wilson, W.; Kingsman, S.M.; Kingsman, A.J.
Nature 313, 243-246, 1985
A:Title: A retrovirus-like strategy for expression of a fusion protein encoded by yea
A:Reference number: A21856; MUID:85111117
A:Accession: S05890
A:Molecule type: DNA
A:Residues: 1-6, '1', 8-10, 'K', 12-53 <MEL>
A:Cross-references: EMBL:X01736; NID:94708; PIDN:CAA25874.1; PID:g1326020
C:Genetics:
A:Mobile element: retrotransposon Ty912
C:Superfamily: Tyb protein

Query Match Similarity 26.6%; Score 55.5; DB 2; Length 1278;
Best Local Similarity 43.8%; Pred. No. 35;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

QY 3 SYGNSADAPYVGAKIGOV---DAKQINGKNT 31
      |||| |::|| : | | | |
Db 1135 SYGN---QPYKKSQIGNIPLNGKVIIGKST 1162

RESULT 6
S52481
Tyb protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty1
N:Alternate names: protein YMO109w
C:Species: Saccharomyces cerevisiae
C:Date: 08-May-1995 #sequence_revision 21-Jul-1995 #text_change 20-Jun-2000
A:Accession: S52481
R:Connor, R.; Churcher, C.
submitted to the EMBL Data Library, February 1995
A:Reference number: S52478
A:Accession: S52481
A:Molecule type: DNA
A:Residues: 1-1328 <CON>
A:Cross-references: EMBL:Z48430; NID:9683664; PIDN:CAA88330.1; PID:g1326005
C:Genetics:
A:Map position: 13L
A:Mobile element: retrotransposon Ty1
C:Superfamily: Tyb protein

Query Match Similarity 26.6%; Score 55.5; DB 2; Length 1328;
Best Local Similarity 43.8%; Pred. No. 37;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

QY 3 SYGNSADAPYVGAKIGOV---DAKQINGKNT 31
      |||| |::|| : | | | |
Db 1185 SYGN---QPYKKSQIGNIYLNGKVIIGKST 1212

```

```
RESULT 7
SY2894
TYB protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty1-9532
N:Alternate names: protein YM9532.10c; protein YMR045c
C:Species: Saccharomyces cerevisiae
C:Date: 08-May-1995 #sequence_revision 01-Sep-1995 #text_change 20-Jun-2000
C:Accession: S52894
R:Odell, C.; Bowman, S.
submitted to the EMBL Data Library, February 1995
A:Reference number: S52885
A:Accession: S52894
A:Molecule type: DNA
A:Residues: 1-1328 <ODE>
A:Cross-references: EMBL:Z48502; NID:g695715; PIDN:CAA88411.1; PID:g1326011
C:Genetics:
A:Map position: 13R
A:Mobile element: retrotransposon Ty1-9532
C:Superfamily: Tyb protein

Query Match          26.6%; Score 55.5; DB 2; Length 1328;
Best Local Similarity 43.8%; Pred. No. 37;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

QY 3 SYGNSADAPYVGAKIGQV---DAKQINGKNT 31
||||| ||| :||| : |||||
Db 1185 SYGN---QPYKSQLGNIFLLNGKVIIGKST 1212

RESULT 8
SY0948
TYB protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty1
N:Alternate names: protein YM9827.08; protein YML045w
C:Species: Saccharomyces cerevisiae
C:Date: 10-Feb-1995 #sequence_revision 12-May-1995 #text_change 20-Jun-2000
C:Accession: S50948
R:Odell, C.; Bowman, S.
submitted to the EMBL Data Library, January 1995
A:Reference number: S50941
A:Accession: S50948
A:Molecule type: DNA
A:Residues: 1-1328 <ODE>
A:Cross-references: EMBL:Z47816; NID:g642303; PIDN:CAA87830.1; PID:g1326015
C:Genetics:
A:Map position: 13L
A:Mobile element: retrotransposon Ty1
C:Superfamily: Tyb protein

Query Match          26.6%; Score 55.5; DB 2; Length 1328;
Best Local Similarity 43.8%; Pred. No. 37;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

QY 3 SYGNSADAPYVGAKIGQV---DAKQINGKNT 31
||||| ||| :||| : |||||
Db 1185 SYGN---QPYKSQLGNIFLLNGKVIIGKST 1212

RESULT 9
B28097
TYB protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty1-H3 (fragment)
C:Species: Saccharomyces cerevisiae
C:Date: 03-Nov-1988 #sequence_revision 25-Apr-1997 #text_change 25-Apr-1997
C:Accession: B28097
R:Boeke, J.D.; Eichinger, D.; Gastrillon, D.; Fink, G.R.
Mol. Cell. Biol. 8, 1432-1442, 1988
A:Title: The Saccharomyces cerevisiae genome contains functional and nonfunctional copies
A:Reference number: A28097; MUID:88246410
A:Accession: B28097
A:Molecule type: DNA
A:Residues: 1-1328 <BOE>
```

```
A:Cross-references: EMBL:M18706
C:Genetics:
A:Mobile element: retrotransposon Ty1-H3
C:Superfamily: Tyb protein

Query Match          26.6%; Score 55.5; DB 2; Length 1328;
Best Local Similarity 43.8%; Pred. No. 37;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

QY 3 SYGNSADAPYVGAKIGQV---DAKQINGKNT 31
||||| ||| :||| : |||||
Db 1185 SYGN---QPYKSQLGNIFLLNGKVIIGKST 1212

RESULT 10
B23496
TYB protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty1-17
C:Species: Saccharomyces cerevisiae
C:Date: 20-Aug-1987 #sequence_revision 20-Aug-1987 #text_change 25-Apr-1997
C:Accession: B23496
R:Warrington, J.R.; Waring, R.B.; Newlon, C.S.; Indge, K.J.; Oliver, S.G.
Nucleic Acids Res. 13, 6679-6693, 1985
A:Title: Nucleotide sequence characterization of Ty 1-17, a class II transposon from
A:Reference number: A93591; MUID:86041864
A:Accession: B23496
A:Molecule type: DNA
A:Residues: 1-1348 <MAR>
A:Cross-references: EMBL:X03840; EMBL:J01333; EMBL:K00633; EMBL:X02991
C:Genetics:
A:Mobile element: retrotransposon Ty1-17
C:Superfamily: Tyb protein

Query Match          26.6%; Score 55.5; DB 2; Length 1348;
Best Local Similarity 43.8%; Pred. No. 37;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

QY 3 SYGNSADAPYVGAKIGQV---DAKQINGKNT 31
||||| ||| :||| : |||||
Db 1205 SYGN---QPYKSQLGNIFLLNGKVIIGKST 1232

RESULT 11
S69972
TYB protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty1.N
N:Alternate names: protein N2453
C:Species: Saccharomyces cerevisiae
C:Date: 09-Mar-1996 #sequence_revision 06-Sep-1996 #text_change 20-Jun-2000
C:Accession: S69972
R:Bergez, P.; Dolignon, F.; Crouzet, M.
submitted to the Protein Sequence Database, April 1996
A:Reference number: S62975
A:Accession: S69972
A:Molecule type: DNA
A:Residues: 1-1749 <BER>
A:Cross-references: EMBL:Z71330; NID:g1301918; PIDN:CAA95924.1; PID:g1301920
A:Note: biosynthesis of this protein involves a +1 frameshift in the codon for residu
C:Genetics:
A:Map position: 14L
A:Mobile element: retrotransposon Ty1.N
C:Superfamily: Tyb protein

Query Match          26.6%; Score 55.5; DB 2; Length 1749;
Best Local Similarity 43.8%; Pred. No. 48;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

QY 3 SYGNSADAPYVGAKIGQV---DAKQINGKNT 31
||||| ||| :||| : |||||
Db 1606 SYGN---QPYKSQLGNIFLLNGKVIIGKST 1633
```

```
RESULT 12
S50641
TYB protein YER138c - yeast (Saccharomyces cerevisiae) retrotransposon Ty
C:Species: Saccharomyces cerevisiae
C:Date: 28-May-1993 #sequence_revision 24-Feb-1995 #text_change 26-Aug-1999
C:Accession: S50641
R:Dietrich, F.S.
submitted to the EMBL Data Library, December 1994
A:Description: The sequence of S. cerevisiae cosmid 8229, 9115, 9132, 9981, and lambda
A:Reference number: S50428
A:Accession: S50641
A:Molecule type: DNA
A:Residues: 1-1755 <DE>
A:Cross-references: EMBL:U18917; NID:g603377; PIDN:AAB64665.1; PID:g603378; MIPS:YER138c
C:Genetics:
A:Map position: 5R
C:Superfamily: Tyb protein

Query Match          26.6%; Score 55.5; DB 2; Length 1755;
Best Local Similarity 43.8%; Pred. No. 49;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY      3 SYGNSADAQPYVGAKIGOV---DAKQINGKNT 31
      |||| |||::||:|:| |||||
Db      1612 SYGN-----QPYKKSQIGNITLNGKVIQKST 1639

RESULT 13
S50663
TYB protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty1.E
N:Alternate names: protein YER160c
C:Species: Saccharomyces cerevisiae
C:Date: 28-May-1993 #sequence_revision 24-Feb-1995 #text_change 26-Aug-1999
C:Accession: S50663; S30812; S53556
R:Dietrich, F.S.
submitted to the EMBL Data Library, December 1994
A:Description: The sequence of S. cerevisiae cosmid 8229, 9115, 9132, 9981, and lambda
A:Reference number: S50428
A:Accession: S50663
A:Molecule type: DNA
A:Residues: 1-1755 <DE>
A:Cross-references: EMBL:U18917; NID:g603377; PIDN:AAB64687.1; PID:g603400
R:Mulligan, J.T.; Dietrich, F.S.; Hennessey, K.M.; Sehl, P.; Komp, C.; Wei, Y.; Taylor,
submitted to the EMBL Data Library, February 1993
A:Reference number: S30812
A:Accession: S30812
A:Molecule type: DNA
A:Residues: 1-230 <MU>
A:Cross-references: EMBL:L10718
C:Genetics:
A:Map position: 5R
A:Mobile element: retrotransposon Ty1.E
C:Superfamily: Tyb protein

Query Match          26.6%; Score 55.5; DB 2; Length 1755;
Best Local Similarity 43.8%; Pred. No. 49;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY      3 SYGNSADAQPYVGAKIGOV---DAKQINGKNT 31
      |||| |||::||:|:| |||||
Db      1612 SYGN-----QPYKKSQIGNITLNGKVIQKST 1639

RESULT 14
S57047
TYB protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty1.JR
N:Alternate names: protein J1570; protein YUR029w
C:Species: Saccharomyces cerevisiae
C:Date: 08-Jul-1995 #sequence_revision 08-Sep-1995 #text_change 17-Mar-1999
C:Accession: S57047; S60512
R:Zagulski, M.; Babinska, B.; Gromadka, R.; Migdalski, A.; Rytko, J.; Sulicka, J.; Herba
```

```
submitted to the Protein Sequence Database, September 1995
A:Reference number: S57040
A:Accession: S57047
A:Molecule type: DNA
A:Residues: 1-1755 <HU>
A:Cross-references: EMBL:Z49528
A:Note: biosynthesis of this protein involves a +1 frameshift in the codon for residu
R:Zagulski, M.; Babinska, B.; Gromadka, R.; Migdalski, A.; Rytko, J.; Sulicka, J.; He
A:Title: The sequence of 24.3 kb from chromosome X reveals five complete open reading
A:Reference number: S60503; MUID:96109930
A:Accession: S60512
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 'T1STFT', 435-1755 <ZAG>
A:Cross-references: EMBL:X87297
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1995
A:Note: the difference at the amino end is due to translation from an incorrect readi
C:Genetics:
A:Map position: 10R
A:Mobile element: retrotransposon Ty1.JR
C:Superfamily: Tyb protein

Query Match          26.6%; Score 55.5; DB 2; Length 1755;
Best Local Similarity 43.8%; Pred. No. 49;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;
```

```
OY      3 SYGNSADAQPYVGAKIGOV---DAKQINGKNT 31
      |||| |||::||:|:| |||||
Db      1612 SYGN-----QPYKKSQIGNITLNGKVIQKST 1639

RESULT 15
S57045
TYB protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty1.JR
N:Alternate names: protein J1560; protein YUR027w
C:Species: Saccharomyces cerevisiae
C:Date: 08-Jul-1995 #sequence_revision 08-Sep-1995 #text_change 17-Mar-1999
R:Zagulski, M.; Babinska, B.; Gromadka, R.; Migdalski, A.; Rytko, J.; Sulicka, J.; He
submitted to the Protein Sequence Database, September 1995
A:Reference number: S57040
A:Accession: S57045
A:Molecule type: DNA
A:Residues: 1-1755 <HU>
A:Cross-references: EMBL:Z49526
A:Note: biosynthesis of this protein involves a +1 frameshift in the codon for residu
R:Zagulski, M.; Babinska, B.; Gromadka, R.; Migdalski, A.; Rytko, J.; Sulicka, J.; He
A:Title: The sequence of 24.3 kb from chromosome X reveals five complete open reading
A:Reference number: S60503; MUID:96109930
A:Accession: S60511
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 'T1STFT', 435-1755 <ZAG>
A:Cross-references: EMBL:X87297
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1995
A:Note: the difference at the amino end is due to translation from an incorrect readi
C:Genetics:
A:Map position: 10R
A:Mobile element: retrotransposon Ty1.JR
C:Superfamily: Tyb protein

Query Match          26.6%; Score 55.5; DB 2; Length 1755;
Best Local Similarity 43.8%; Pred. No. 49;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;
```

Tue Mar 6 16:25:07 2001

us-09-164-714-1.rpt

Page 5

Search completed: March 6, 2001, 14:37:44
Job time: 117 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: March 6, 2001, 14:37:12 ; Search time 14.8 Seconds
(without alignments)
87,281 Million cell updates/sec

Title: US-09-164-714-1

Perfect score: 209
Sequence: 1 A1SYGNSADAPYVGAKIGQVDAKQINGKNTAYGIVAGYN 40

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	55.5	26.6	289	1	P39767 rhodopseudo
2	55.5	26.6	1328	1	Q03434 saccharomyc
3	55.5	26.6	1328	1	Q04711 saccharomyc
4	55.5	26.6	1328	1	YMT5_YEAST
5	55.5	26.6	1328	1	YMD0_YEAST
6	55.5	26.6	1347	1	YCB9_YEAST
7	55.5	26.6	1755	1	Y277_YEAST
8	55.5	26.6	1755	1	Y299_YEAST
9	54	25.8	1092	1	DPOD_DROME
10	53.5	25.6	493	1	MOO_MYCTU
11	53	25.4	1249	1	TPP2_RAT
12	51	24.4	199	1	UREG_HELPJ
13	51	24.4	199	1	UREG_HELPJ
14	50.5	24.2	250	1	YHIO_NELGO
15	50	23.9	121	1	OPA_HAEIN
16	50	23.9	325	1	Y05F_BPT4
17	49.5	23.7	77	1	PURX_SCHPO
18	49.5	23.7	449	1	COMB_STRPN
19	49.5	23.7	584	1	FRAC_BORPE
20	49	23.4	357	1	ALF_SP10L
21	49	23.4	358	1	ALF_ARATH
22	49	23.4	540	1	CHT1_RH10L
23	49	23.4	959	1	SYAC_SCHPO
24	49	23.4	1116	1	SLPH_BACBR
25	49	23.4	1262	1	TPP2_MOUSE
26	48.5	23.2	453	1	KICH_RAT
27	48.5	23.2	1156	1	G9A_BACNG
28	48	23.0	131	1	R56E_METJA
29	48	23.0	204	1	UREG_STRSL
30	48	23.0	206	1	UREG_STRSL
31	48	23.0	313	1	BGL2_YEAST
32	48	23.0	518	1	FUS_MOUSE
33	48	23.0	1088	1	DPOD_SOYBN

34	48	23.0	1242	1	RP02_ASFB7
35	48	23.0	1842	1	PAS2_SCHPO
36	47.5	22.7	472	1	PEB_ESCFE
37	47.5	22.7	629	1	VE1_HPV31
38	47.5	22.7	1133	1	LOMM_YEAST
39	47	22.5	241	1	CAMT_PETCR
40	47	22.5	242	1	CAMT_VITVI
41	47	22.5	351	1	MS52_YEAST
42	47	22.5	360	1	PGS2_PIG
43	47	22.5	750	1	YKS7_YEAST
44	47	22.5	788	1	PBS2_YEAST
45	46.5	22.2	329	1	SYFA_HAEIN

ALIGNMENTS

RESULT	ID	PORT_RHOBL	STANDARD	PRT	289 AA.
AC	P39767				
DT	01-FEB-1995	(Rel. 31, Created)			
DT	01-FEB-1995	(Rel. 31, Last sequence update)			
DT	15-JUL-1999	(Rel. 38, Last annotation update)			
DE	PORIN.				
GN	OPMA.				
OS	Rhodopseudomonas blattica.				
OC	Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group; Rhodobacter.				
RN	[1]				
RP	X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).				
RX	MEDLINE=94191532; PubMed=8142898;				
RA	Kreusch A., Neuhueser A., Schiltz E., Weckesser J., Schulz G.E.;				
RT	Structure of the membrane channel porin from Rhodopseudomonas				
RT	blattica at 2.0-A resolution."				
RL	Protein Sci. 3:58-63(1994).				
RN	[2]				
RP	X-RAY CRYSTALLOGRAPHY (1.96 ANGSTROMS).				
RX	MEDLINE=95055730; PubMed=7525973;				
RA	Kreusch A., Schulz G.E.;				
RT	Refined structure of the porin from Rhodopseudomonas blattica.				
RT	Comparison with the porin from Rhodobacter capsulatus."				
RL	J. Mol. Biol. 243:891-905(1994).				
RN	[3]				
RP	X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF MUTANTS.				
RX	MEDLINE=96848035; PubMed=9684893;				
RA	Schmid B., Weveraud L., Kromer M., Schulz G.E.;				
RT	"Porin mutants with new channel properties."				
RT	Protein Sci. 7:1603-1611(1998).				
CC	- FUNCTION: FORMS CHANNELS THAT ALLOW THE PASSIVE DIFFUSION OF SMALL				
CC	HYDROPHILIC SOLUTIONS UP TO AN EXCLUSION LIMIT OF ABOUT 0.6 KDA.				
CC	- SUBUNIT: HOMOTRIMER.				
CC	- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.				
DR	PIR: S38806; S38806.				
DR	PDB: 1PRN; 24-OCT-94.				
DR	PDB: 2PRN; 13-JAN-99.				
DR	PDB: 3PRN; 12-AUG-98.				
DR	PDB: 5PRN; 12-AUG-98.				
DR	PDB: 6PRN; 12-AUG-98.				
DR	PDB: 7PRN; 12-AUG-98.				
DR	PDB: 8PRN; 12-AUG-98.				
DR	PDB: 1BH3; 12-AUG-98.				
KW	Outer membrane; Transmembrane; Porin; 3D-structure.				
SEQ	SEQUENCE 289 AA; 30597 MW; 0825209803A1044C CRC64;				

Query Match 26.6%; Score 55.5; DB 1; Length 289;

Best Local Similarity 41.7%; Pred. No. 3.7;

Matches 15; Conservative 3; Mismatches 15; Indels 3; Gaps 1;

QY 4 YGNSADAPYVGAKIGQVDAKQINGKNTAYGIVAGY 39
DB 227 YGNVAFGATTVRAVYVSDIDRA--GADTAYGIGADY 259

```
RESULT 2
YMD9_YEAST STANDARD: PRT; 1328 AA.
AC 003434;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE TRANSPONSON TY1 PROTEIN B.
FT TY1B OR YML039W OR YM8054.04.
GN Saccharomyces cerevisiae (Baker's yeast).
OS Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA Connor R., Churcher C., Barrell B.G., Rajandream M.A.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY U23.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; 248430; CAA88330.1; -.
CC SGD; S0004503; YML039W.
CC INTERPRO: IPR001969; -.
CC PROSITE; PS00141; ASP_PROTEASE; 1.
CC Transposable element: Hypothetical protein; Hydrolase;
CC Aspartyl protease; ATP-binding.
CC ACT_SITE 34 34 PROTEASE (BY SIMILARITY).
CC NP_BIND 1204 1211 ATP (POTENTIAL).
CC FT SEQUENCE 1328 AA; 151036 MW; C7D14E1AA675E93 CRC64;
CC SQ

Query Match 26.6%; Score 55.5; DB 1; Length 1328;
Best Local Similarity 43.8%; Pred. No. 17;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY 3 SYGNASDAQPYVGAKIGOV---DAKQINGKNT 31
Db 1185 SYGN---QPYKSGQIGNIYLNGKVGIGKST 1212

RESULT 3
YME4_YEAST STANDARD: PRT; 1328 AA.
AC 004711;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE TRANSPONSON TY1 PROTEIN B.
FT TY1B OR YML044W OR YM9827.08.
GN Saccharomyces cerevisiae (Baker's yeast).
OS Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA Odell C., Bowman S., Barrell B.G., Rajandream M.A.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY U23.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; 248502; CAA88411.1; -.
CC SGD; S0004648; YMR045C.
CC INTERPRO: IPR001969; -.
CC PROSITE; PS00141; ASP_PROTEASE; 1.
CC Transposable element: Hypothetical protein; Hydrolase;
CC Aspartyl protease; ATP-binding.
CC ACT_SITE 34 34 PROTEASE (BY SIMILARITY).
CC NP_BIND 1204 1211 ATP (POTENTIAL).
CC FT SEQUENCE 1328 AA; 151187 MW; AA19E50B62BA43F95 CRC64;
CC SQ

Query Match 26.6%; Score 55.5; DB 1; Length 1328;
Best Local Similarity 43.8%; Pred. No. 17;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY 3 SYGNASDAQPYVGAKIGOV---DAKQINGKNT 31
Db 1185 SYGN---QPYKSGQIGNIYLNGKVGIGKST 1212

RESULT 5
YMU0_YEAST STANDARD: PRT; 1328 AA.
AC 004670;
```

```
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; 247816; CAA87830.1; -.
CC INTERPRO: IPR001969; -.
CC PROSITE; PS00141; ASP_PROTEASE; 1.
CC Transposable element: Hypothetical protein; Hydrolase;
CC Aspartyl protease; ATP-binding.
CC ACT_SITE 34 34 PROTEASE (BY SIMILARITY).
CC NP_BIND 1204 1211 ATP (POTENTIAL).
CC FT SEQUENCE 1328 AA; 150945 MW; E9B964CD76CD281F CRC64;
CC SQ

Query Match 26.6%; Score 55.5; DB 1; Length 1328;
Best Local Similarity 43.8%; Pred. No. 17;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY 3 SYGNASDAQPYVGAKIGOV---DAKQINGKNT 31
Db 1185 SYGN---QPYKSGQIGNIYLNGKVGIGKST 1212

RESULT 4
YMT5_YEAST STANDARD: PRT; 1328 AA.
AC 004214;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE TRANSPONSON TY1 PROTEIN B.
GN TY1B OR YMR045C OR YM9532.10C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA Odell C., Bowman S., Barrell B.G., Rajandream M.A.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY U23.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; 248502; CAA88411.1; -.
CC SGD; S0004648; YMR045C.
CC INTERPRO: IPR001969; -.
CC PROSITE; PS00141; ASP_PROTEASE; 1.
CC Transposable element: Hypothetical protein; Hydrolase;
CC Aspartyl protease; ATP-binding.
CC ACT_SITE 34 34 PROTEASE (BY SIMILARITY).
CC NP_BIND 1204 1211 ATP (POTENTIAL).
CC FT SEQUENCE 1328 AA; 151187 MW; AA19E50B62BA43F95 CRC64;
CC SQ

Query Match 26.6%; Score 55.5; DB 1; Length 1328;
Best Local Similarity 43.8%; Pred. No. 17;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY 3 SYGNASDAQPYVGAKIGOV---DAKQINGKNT 31
Db 1185 SYGN---QPYKSGQIGNIYLNGKVGIGKST 1212

RESULT 5
YMU0_YEAST STANDARD: PRT; 1328 AA.
AC 004670;
```

DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE TRANSDON TY1 PROTEIN B.
 GN TY1B OR YMR050C OR YMR9796.03C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 CC Saccharomycetaceae; Saccharomyces.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C / AB972.
 RA Devlin K., Churcher C.M., Barrell B.G., Rajandream M.A.;
 RA Submitted (MAY-1995) to the EMBL/Genbank/DBJ databases.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY U23.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: Z49703; CAAB9760.1; -
 CC SGD: S0004653; YMR050C.
 DR INTERPRO: IPR001969; -
 DR PROSITE: PS00141; ASP_PROTEASE: 1.
 KW Transposable element; Hypothetical protein; Hydrolase;
 KM Aspartyl protease; ATP-binding
 FT ACT_SITE 34 34 PROTEASE (BY SIMILARITY).
 FT NP_BIND 1204 1211 ATP (POTENTIAL).
 SQ SEQUENCE 1328 AA; 150967 MW; DACD7A471697DIDD CRC64;

Query Match 26.6%; Score 55.5; DB 1; Length 1328;
 Best Local Similarity 43.8%; Pred. No. 17;
 Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY 3 SYGNSADQPYGAKIGOV---DAKQINGKNT 31
 ||||| ||| :||| :| |||||
 DB 1185 SYGN---QPYKSSQIGNIYLNGKVIQKST 1212

RESULT 6
 YCB9_YEAST STANDARD; PRT; 1347 AA.
 AC P25384;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE TRANSDON TY1-17 PROTEIN B.
 GN TY1B OR YCL019W OR YCL19W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 CC Saccharomycetaceae; Saccharomyces.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE-86205247; PubMed=3010239;
 RA Warrington J.R., Anwar R., Newton C.S., Waring R.B., Davies R.W.,
 RA Indge K.J., Oliver S.G.;
 RA "A 'hot-spot' for Ty transposition on the left arm of yeast
 RT chromosome III."
 RL Nucleic Acids Res. 14:3475-3485(1986).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC MEDLINE-86041864; PubMed=2997719;
 RA Warrington J.R., Waring R.B., Newton C.S., Indge K.J., Oliver S.G.;
 RT "Nucleotide sequence characterization of Ty 1-17, a class II
 RT transposon from yeast."
 RN Nucleic Acids Res. 13:6679-6693(1985).
 RP [3]
 RP SEQUENCE FROM N.A.
 RA Oliver S.G., Anwar R., Brown A., Gent M.E., Indge K.J., James C.M.,

RA Stateva L.I.;
 RL Submitted (MAR-1992) to the EMBL/Genbank/DBJ databases.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY U23.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: X03840; CAA27458.1; -
 CC EMBL: X59720; E264443; -
 DR PIR: B23496; B23496.
 DR PIR: S19345; S19345.
 DR MEROPS: A11.003; -
 DR SGD: S0000524; YCL019W.
 KW Transposable element; Hypothetical protein; Hydrolase;
 KM Aspartyl protease; ATP-binding.
 FT NP_BIND 1223 1230 ATP (POTENTIAL).
 SQ SEQUENCE 1347 AA; 154069 MW; AD3660C5E78282HF CRC64;

Query Match 26.6%; Score 55.5; DB 1; Length 1347;
 Best Local Similarity 43.8%; Pred. No. 17;
 Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY 3 SYGNSADQPYGAKIGOV---DAKQINGKNT 31
 ||||| ||| :||| :| |||||
 DB 1204 SYGN---QPYKSSQIGNIYLNGKVIQKST 1231

RESULT 7
 YJZ7_YEAST STANDARD; PRT; 1755 AA.
 AC P47098; P47194;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE TRANSDON TY1 PROTEIN B.
 GN TY1B OR YJR027W OR J1560.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 CC Saccharomycetaceae; Saccharomyces.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE-96109930; PubMed=8619316;
 RA Zagulski M., Babincka B., Gromadka R., Migdalski A., Rytko J.,
 RA Sulicka J., Herbert C.J.;
 RT "The sequence of 24.3 kb from chromosome X reveals five complete open
 RT reading frames, all of which correspond to new genes, and a tandem
 RT insertion of a Ty1 transposon."
 RL Yeast 11:1179-1186(1995).
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY U23.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: Z49526; CAA89553.1; -
 CC MEROPS: A11.003; -
 DR SGD: S0003788; YJR027W.
 DR INTERPRO: IPR001042; -
 DR PFAM: PF01021; TYA: 1.
 DR PROSITE: PS00141; ASP_PROTEASE: 1.
 KW Transposable element; Hypothetical protein; Hydrolase;
 KM Aspartyl protease; ATP-binding.
 FT ACT_SITE 461 PROTEASE (BY SIMILARITY).
 FT ACT_SITE 461

KW Nickel: ATP-binding. 15 ATP (POTENTIAL).
FT NP_BIND 8
SO SEQUENCE 199 AA; 21941 MW; 6959FFEB6A3D3C1 CRC64;
Matches 11; Conservative 6; Mismatches 7; Indels 2; Gaps 1;
Query Match 24.4%; Score 51; DB 1; Length 199;
Best Local Similarity 42.3%; Pred. No. 10;
Matches 11; Conservative 6; Mismatches 7; Indels 2; Gaps 1;
OY 6 NSADAQPYVGA--KIGOVDAKOINCK 29
DB 145 NKIDLAPYVGADLKVMERSKMRGE 170
RESULT 13
UREG_HELPY STANDARD; PRT; 199 AA.
AC 009066;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE UREASE ACCESSORY PROTEIN UREG.
GN UREG OR HP0068.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-85P;
RX MEDLINE-92210488; PubMed-1313413;
RA Cussac V., Ferrero R.L., Labigne A.;
RT "Expression of Helicobacter pylori urease genes in Escherichia coli
grown under nitrogen-limiting conditions.";
RL J. Bacteriol. 174:2466-2473(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-26695 / ATCC 700392;
RX MEDLINE-97394467; PubMed-9252185;
RA Tomb J.-F., White O., Kervilange A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Fitzgerald D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Maithey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
pylori.";
RL Nature 388:539-547(1997).
CC -1- FUNCTION: PROBABLY FACILITATING NICKEL INCORPORATION.
CC -1- SIMILARITY: BELONGS TO THE UREG FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M84338; AAA25025.1; -;
DR EMBL; A24198; CAA01726.1; -;
DR EMBL; AE000528; AAD07131.1; -;
DR PIR; D41834; D41834.
DR TIGR; HP0068; -;
DR INTERPRO; IPR002894; -;
DR PFAM; PF01495; Hypr_ureg; 1.
KW Nickel: ATP-binding.
FT NP_BIND 8
FT CONFLICT 102 102 D -> S (IN REF. 1).
FT CONFLICT 167 176 MRGKPFET -> IAAKSPLELP (IN REF. 1).
SO SEQUENCE 199 AA; 21955 MW; A23E448EB6A208B1 CRC64;

Query Match 24.4%; Score 51; DB 1; Length 199;
Best Local Similarity 42.3%; Pred. No. 10;
Matches 11; Conservative 6; Mismatches 7; Indels 2; Gaps 1;
OY 6 NSADAQPYVGA--KIGOVDAKOINCK 29
DB 145 NKIDLAPYVGADLKVMERSKMRGE 170
RESULT 14
YHIQ_NEIGO STANDARD; PRT; 250 AA.
AC P72077;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHETICAL 27.3 KDA PROTEIN.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-M511;
RA Carrick C.S., Pyfe J.A.M., Davies J.K.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: STRONG, TO E.COLI YHIQ AND H.INFLUENZAE H10949.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U65994; AAC82508.1; -;
DR Hypothetical protein.
KW SEQUENCE 250 AA; 27316 MW; FC749D6B91985763 CRC64;
SO
Query Match 24.2%; Score 50.5; DB 1; Length 250;
Best Local Similarity 36.7%; Pred. No. 15;
Matches 11; Conservative 6; Mismatches 12; Indels 1; Gaps 1;
OY 11 OYVGAKI-GOVDAKOINKMTAVGIVGY 39
DB 216 RPRGEHLAQAAPAYQYTGKSTREVDYLPY 245
RESULT 15
ID OPA_HAEIN STANDARD; PRT; 121 AA.
AC P45088;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE OPA PROTEIN.
GN OPA OR H1174.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RD / KW20;
RX MEDLINE-95350630; PubMed-7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kervilange A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fitchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,

THIS PAGE BLANK (USPTO)

Db 77 Y 77

RESULT 2

09XTP9 PRELIMINARY: PRT: 380 AA.

AC 09XTP9; 01-NOV-1999 (TREMblrel. 12, Created)

DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)

DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)

DE ZK1321.4 PROTEIN.

GN ZK1321.4.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Pelodierinae; Caenorhabditis.

OX NCBI_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=94150718; PubMed=7906398;

RA Gardner A.;

RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

RT elegans.";

RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=94150718; PubMed=7906398;

RA Wilson R., Alnscough R., Anderson K., Baynes C., Berks M., Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A., Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L., Jones M., Kershaw J., Kirsten J., Laister N., Latreille P., Lightning J., Lloyd C., McMurtry A., Mortimore B., O'Callaghan M., Parsons J., Percy C., Rifken L., Koopra A., Saunders D., Showkeen R., Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J., Thelery-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterson R., Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;

RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

RT elegans.";

RL Nature 368:32-38(1994).

RX EMBL; Z48717; CA88612.1; JOINED.

DR EMBL; Z48584; CA88612.1; JOINED.

DR EMBL; Z48584; CA88612.1; JOINED.

DR EMBL; Z48717; CA88612.1; JOINED.

SO SEQUENCE 380 AA; 40259 MW; DC749A7E6A488BF3 CRC64;

Query Match 28.5%; Score 59.5; DB 5; Length 380;

Best Local Similarity 38.6%; Pred. No. 6.3;

Matches 17; Conservative 2; Mismatches 16; Indels 9; Gaps 2;

OY 3 SYGNSADAPYVGAKIGVDKQIN-----GKNTAYGIYAG 38

Db 224 SYNGATSGPY-GAGSGGTPLNQMTFINTSPAPGANGAYGACG 266

RESULT 3

09KS75 PRELIMINARY: PRT: 160 AA.

AC 09KS75; 01-OCT-2000 (TREMblrel. 15, Created)

DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)

DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)

DE HYPOTHETICAL PROTEIN VC1384.

GN VC1384.

OS Vibrio cholerae.

OC Bacteria; Proteobacteria; gamma subdivision: Vibrionaceae; Vibrio.

OX NCBI_TaxID=666;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=EL TOR N16961 / SEROTYPE O1;

RX MEDLINE=20406833; PubMed=10952301;

RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,

RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A., Gill S.R., Nelson K.E., Read T.D., Tellein H., Richardson D., Ermolaeva M.D., Yamathavan J., Bass S., Qin H., Dragol I., Sellers P., McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O., Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C., Fraser C.M.;

RT "DNA sequence of both chromosomes of the cholera pathogen vibrio

RT cholerae.";

RL Nature 406:477-483(2000).

DR EMBL; AE004217; AAF94542.1; -.

DR TIGR; VC1384; -.

KW Hypothetical protein.

SO SEQUENCE 160 AA; 16985 MW; 7E47CED047CE409D CRC64;

Query Match 28.2%; Score 59; DB 2; Length 160;

Best Local Similarity 45.2%; Pred. No. 2.7;

Matches 14; Conservative 6; Mismatches 9; Indels 2; Gaps 1;

OY 8 ADAOPYVGAKIGVDKQINCKNTAYGIYAG 38

Db 19 ADSWIYGASVGSQSDYEGKHG--TAYSVHAG 47

RESULT 4

051841 PRELIMINARY: PRT: 346 AA.

AC 051841; 01-JUN-1998 (TREMblrel. 06, Created)

DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)

DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)

DE OUTER MEMBRANE PROTEIN 34 PRECURSOR.

GN OMP34.

OS Actinobacillus actinomycetemcomitans

OS (Haemophilus actinomycetemcomitans).

OC Bacteria; Proteobacteria; gamma subdivision: Pasteurellaceae;

OC Actinobacillus.

OX NCBI_TaxID=714;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=NCTC 9710;

RX MEDLINE=98084499; PubMed=9423883;

RA White P.A., Nair S.P., Kim M.J., Wilson M., Henderson B.;

RT "Molecular characterization of an outer membrane protein of

RT Actinobacillus actinomycetemcomitans belonging to the Ompa family.";

RL Infect. Immun. 66:369-372(1998).

DR EMBL; AF005079; AAC00068.1; -.

DR HSSP; P02934; 10JP.

DR INTERPRO; IPR000498; -.

DR INTERPRO; IPR01145; -.

DR PFAM; PF00691; OmpA; 1.

DR PFAM; PF01389; OmpA_membrane; 1.

DR PRINTS; PRO1021; OMPADOMAIN.

DR PROSITE; PS01068; OMPA; 1.

DR PRODOM; PD000930; -; 1.

KW Signal.

FT SIGNAL.

SO SEQUENCE 346 AA; 36905 MW; 68FFA4BC8B0F1819 CRC64;

Query Match 26.8%; Score 56; DB 2; Length 346;

Best Local Similarity 35.4%; Pred. No. 16;

Matches 17; Conservative 4; Mismatches 11; Indels 16; Gaps 3;

OY 8 ADAOPYVGAKIGVDKQINCKNTAYGIYAG 39

Db 19 AQAAPQANTFYAGAKAGWASSHHGLNDFKQKGVISINRSEAYGVGGY 66

RESULT 5

09S5J9 PRELIMINARY: PRT: 346 AA.

AC 09S5J9;

```
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
DE OMP29.
GN OMP34.
OS Actinobacillus actinomycetemcomitans
OS (Haemophilus actinomycetemcomitans).
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Actinobacillus.
OX NCBI_TaxID=714;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC43718;
RA Komatsuzawa H., Kawai T., Wilson M.E., Taubman M.A., Sugai M.,
  Suganaka H.;
RT "Cloning of the gene encoding the Actinobacillus actinomycetemcomitans
RT serotype beta OmpA-like outer membrane protein.";
RL Infect. Immun. 67:942-945(1999).
DR EMBL; AB015936; BAA75215.1; -.
DR INTERPRO: IPR000498; -.
DR INTERPRO: IPR001035; -.
DR INTERPRO: IPR001145; -.
DR INTERPRO: IPR002368; -.
DR PFAM: PF00691; OmpA; 1.
DR PFAM: PF01389; OmpA_membrane; 1.
DR PRINTS: PR01021; OMPADOMAIN.
DR PRINTS: PR01022; OUTERMBRANEA.
DR PRINTS: PR01023; NAFLGMOTY.
DR PROSITE: PS01068; OMPA; 1.
SQ SEQUENCE 346 AA; 36933 MW; 68E9D5DB8B0F1819 CRC64;
```

```
Query Match 26.8%; Score 56; DB 2; Length 346;
Best Local Similarity 35.4%; Pred. No. 16;
Matches 17; Conservative 4; Mismatches 11; Indels 16; Gaps 3;
```

```
QY 8 ADAQP-----YVGAKIGQYDA-----RQ-----INGKNTAYGIYAGY 39
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 19 AQAAPQANTFYAGAKGKGMASHHGLNQPFQKGVSTNRNEAGVGFEGY 66
```

```
RESULT 6
ID 012357 PRELIMINARY; PRT; 767 AA.
AC 012357;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
DE FRAME-SHIFT IN TYB PROBABLY NOT FUNCTIONAL.
GN TY1B.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Rieger M., Mueller-Auer S., Brueckner M., Schaefer M.;
RN Submitted (May-1996) to the EMBL/Genbank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RA MIPS;
RL Submitted (MAY-1996) to the EMBL/Genbank/DBJ databases.
DR EMBL; Z72824; CAA97037.1; -.
DR EMBL; Z72823; CAA97029.1; -.
DR INTERPRO: IPR000194; -.
DR PROSITE: PS00152; ATPASE_ALPHA_BETA; UNKNOWN.1.
SQ SEQUENCE 767 AA; 87951 MW; 5F69264A58605BA5 CRC64;
```

```
Query Match 26.8%; Score 55.5; DB 3; Length 767;
Best Local Similarity 43.8%; Pred. No. 49;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;
```

```
QY 3 SYGNSADAQPYVGAKIGQV---DAKQINGKNT 31
      |||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 624 SYGN----QPYKSQLGNIIFLLNGKVIQCKST 651
```

```
RESULT 7
ID 007163 PRELIMINARY; PRT; 1155 AA.
AC 007163;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE (STRAIN JB84A CONTAINING PLASMID PNN162).
GN TYB.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JB84A;
RX MEDLINE=88246410; PubMed=2837641;
RA Boeke J.D., Etchlinger D., Castillon D., Fink G.R.;
RT "The Saccharomyces cerevisiae genome contains functional and
RT nonfunctional copies of transposon Ty1.";
RL Mol. Cell. Biol. 8:1432-1442(1988).
DR EMBL; M18706; AAA66938.1; -.
DR INTERPRO: IPR001584; -.
DR PFAM: PF00665; rve; 1.
SQ SEQUENCE 1155 AA; 131935 MW; 2076198CD623484C CRC64;
```

```
Query Match 26.6%; Score 55.5; DB 3; Length 1155;
Best Local Similarity 43.8%; Pred. No. 80;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;
```

```
QY 3 SYGNSADAQPYVGAKIGQV---DAKQINGKNT 31
      |||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1012 SYGN----QPYKSQLGNIIFLLNGKVIQCKST 1039
```

```
RESULT 8
ID 013527 PRELIMINARY; PRT; 1196 AA.
AC 013527;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE YAR009CP.
GN YAR009C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=85182712; PubMed=3886659;
RA Drabkin H.J., Rajbandary U.L.;
RT "Attempted expression of a human Initiator tRNA gene in Saccharomyces
RT cerevisiae.";
RL J. Biol. Chem. 260:5596-5602(1985).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=86028187; PubMed=2996783;
RA Hieter P., Pridmore D., Hegemann J.H., Thomas M., Davis R.W.,
RA Philippsen P.;
RT "Functional selection and analysis of yeast centromeric DNA.";
RL Cell 42:913-921(1985).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
```

RX MEDLINE-89128457; PubMed=2644626;
 RA Slater M.R., Craig E.A.;
 RT "The SSA1 and SSA2 genes of the yeast *Saccharomyces cerevisiae*.";
 RL Nucleic Acids Res. 17:805-806(1989).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C;
 RX MEDLINE-90291999; PubMed=2192864;
 RA Heyer W.D., Rao M.R., Erdlie L.F., Kelly T.J., Kolodner R.D.;
 RT "An essential *Saccharomyces cerevisiae* single-stranded DNA binding
 RL protein is homologous to the large subunit of human RPA-A.";
 RN EMBO J. 9:2321-2329(1990).
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C;
 RX MEDLINE-91071607; PubMed=2253888;
 RA White W., Keop L.H., Lamb J., Crowley J.C., Kaback D.B.;
 RT "Molecular cloning of chromosome I DNA from *Saccharomyces cerevisiae*:
 RL isolation, characterization and regulation of the SP07 sporulation
 gene";
 RN Gene 95:65-72(1990).
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C;
 RX MEDLINE-91353080; PubMed=1882551;
 RA Schweitzer B., Philippson P.;
 RT "CUC15, an essential cell cycle gene in *Saccharomyces cerevisiae*,
 RL encodes a protein kinase domain.";
 RN Yeast 7:265-273(1991).
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C;
 RX MEDLINE-92051323; PubMed=1658741;
 RA Davies C.J., Hutchison C.A. III.;
 RT "A directed DNA sequencing strategy based upon Tn3 transposon
 RL mutagenesis: application to the ADE1 locus on *Saccharomyces cerevisiae*
 RN chromosome I.";
 RN Nucleic Acids Res. 19:5731-5738(1991).
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C;
 RX MEDLINE-93066269; PubMed=1279682;
 RA Lefebvre O., Carles C., Conesa C., Swanson R.N., Bouet F., Riva M.,
 RT "Lefebvre O., Carles C., Conesa C., Swanson R.N., Bouet F., Riva M.,
 RL Sentenac A.;
 RT "TRC3: gene encoding the B-block binding subunit of the yeast
 RL transcription factor ITC.";
 RN Proc. Natl. Acad. Sci. U.S.A. 89:10512-10516(1992).
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C;
 RX MEDLINE-93131037; PubMed=8420802;
 RA Hiraga K., Suzuki K., Tsuchiya E., Miyakawa T.;
 RT "Cloning and characterization of the elongation factor EF-1 beta
 RL homologue of *Saccharomyces cerevisiae*. EF-1 beta is essential for
 growth.";
 RN FEBS Lett. 316:165-169(1993).
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C;
 RX MEDLINE-95028152; PubMed=7941740;
 RA Clark M.W., Keng T., Storms R.K., Zhong W., Fortin N., Zeng B.,
 RT "Sequencing of chromosome I of *Saccharomyces cerevisiae*: analysis of
 RL the 42 kbp SP07-CEN1-CUC15 region.";
 RN Yeast 10:535-541(1994).
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C;
 RX MEDLINE-95249563; PubMed=7731988;
 RA Bussey H., Kaback D.B., Zhong W., Vo D.T., Clark M.W., Fortin N.,
 RT Hall J., Ouellette B.F., Keng T., Barton A.B., Su Y., Davies C.K.,
 RA Storms R.K.;
 RT "The nucleotide sequence of chromosome I from *Saccharomyces*

RT *cerevisiae*.";
 RL Proc. Natl. Acad. Sci. U.S.A. 92:3809-3813(1995).
 RN [12]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C;
 RA Jia Y., Cherry J.M.;
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: L22015; AOC04967.1; -.
 DR INTERPRO: IPR001584; -.
 DR Pfam: PF00665; Irf: 1.
 SQ SEQUENCE 1196 AA; 136731 MW; 0A3E8CFC5B12C6A2 CRC64;
 Query Match 26.6%; Score 55.5; DB 3; Length 1196;
 Best Local Similarity 43.8%; Pred. NO. 83;
 Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;
 QY 3 SYGNSADAPYVAKIGOV---DAKQNGKNT 31
 DB 1053 SYGN---QPYKSOIGNYLNGKVGIGKST 1080
 RESULT 9
 ID 003934 PRELIMINARY; PRT; 1285 AA.
 AC 003934;
 DT 01-NOV-1996 (TRENBLREL. 01, Created)
 DF 01-NOV-1996 (TRENBLREL. 01, Last sequence update)
 DT 01-OCT-2000 (TRENBLREL. 15, Last annotation update)
 DE HYPOTHETICAL PROTEIN (FRAGMENT).
 GN TYB.
 OS *Saccharomyces cerevisiae* (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; *Saccharomycetes*; *Saccharomycetales*;
 OC *Saccharomycetaceae*; *Saccharomyces*.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-AB972;
 RA Murphy L., Harris D.;
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-AB972;
 RA Barrell B., Rajandream M.A., Walsh S.V.;
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
 DR EMBL: Z68329; CAA92721.1; -.
 DR INTERPRO: IPR000194; -.
 DR INTERPRO: IPR001584; -.
 DR INTERPRO: IPR001969; -.
 DR Pfam: PF00665; Irf: 1.
 DR PROSITE: PS00141; ASP_PROTEASE; 1.
 DR PROSITE: PS00152; ATPASE_ALPHA_BETA; UNKNOWN_1.
 KW Hypothetical protein; Hydrolase; Aspartyl protease.
 FT NON_TER 1285
 SQ SEQUENCE 1285 AA; 146556 MW; 7E29F748800E52D2 CRC64;
 Query Match 26.6%; Score 55.5; DB 3; Length 1285;
 Best Local Similarity 43.8%; Pred. NO. 91;
 Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;
 QY 3 SYGNSADAPYVAKIGOV---DAKQNGKNT 31
 DB 1203 SYGN---QPYKSOIGNYLNGKVGIGKST 1230
 RESULT 10
 ID 099231 PRELIMINARY; PRT; 1328 AA.
 AC 099231;
 DT 01-NOV-1996 (TRENBLREL. 01, Created)
 DT 01-NOV-1996 (TRENBLREL. 01, Last sequence update)
 DT 01-OCT-2000 (TRENBLREL. 15, Last annotation update)

DE HYPOTHETICAL 151.2 KDA PROTEIN.
GN TYB.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AB972;
RA Oliver K., Shore L., Harris D.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AB972;
RA Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=AB972;
RA Barrell B., Rajandream M.A., Walsh S.V.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL: 268194; CAAG2352.1; -
DR EMBL: 268195; CAAG2360.1; -
DR INTERPRO: IPR001584; -
DR INTERPRO: IPR001969; -
DR PFM: PF00665; IVE: 1.
DR PROSITE: PS00141; ASP_PROTEASE; 1.
KW Hypothetical protein; Hydrolase; Aspartyl protease.
SQ SEQUENCE 1328 AA; 151193 MW; 0E62548690893803 CRC64;

Query Match 26.6%; Score 55.5; DB 3; Length 1328;
Best Local Similarity 43.8%; Pred. No. 94;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY 3 SYGNSADAPYVGAKIGV---DAKOINGKNT 31
Db 1185 SYGN---OPYKSOIGNIFLNGKVGKST 1212

RESULT 11
003855 ID 003855 PRELIMINARY; PRT; 1328 AA.
AC 003855;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, last annotation update)
DE TYB PROTEIN.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AB972;
RA Barrell B., Rajandream M.A.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL: 247746; CAAB7673.1; -
DR INTERPRO: IPR001584; -
DR INTERPRO: IPR001969; -
DR PFM: PF00665; IVE: 1.
DR PROSITE: PS00141; ASP_PROTEASE; 1.
KW Hydrolase; Aspartyl protease.
SQ SEQUENCE 1328 AA; 151007 MW; 6BCHD9FC78B5638 CRC64;

Query Match 26.6%; Score 55.5; DB 3; Length 1328;
Best Local Similarity 43.8%; Pred. No. 94;

Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;
OY 3 SYGNSADAPYVGAKIGV---DAKOINGKNT 31
Db 1185 SYGN---OPYKSOIGNIFLNGKVGKST 1212

RESULT 12
004345 ID 004345 PRELIMINARY; PRT; 1346 AA.
AC 004345;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, last annotation update)
DE HYPOTHETICAL 153.9 KDA PROTEIN.
GN TYB.YD9673.05C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AB972;
RA Connor R., Churcher C.M.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AB972;
RA Barrell B., Rajandream M.A., Walsh S.V.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL: 268196; CAAG2372.1; -
DR INTERPRO: IPR00194; -
DR INTERPRO: IPR001584; -
DR INTERPRO: IPR001969; -
DR PFM: PF00665; IVE: 1.
DR PROSITE: PS00141; ASP_PROTEASE; 1.
DR PROSITE: PS00152; ATPASE_ALPHA_BETA; UNKNOWN.1.
KW Hypothetical protein; Hydrolase; Aspartyl protease.
SQ SEQUENCE 1346 AA; 153889 MW; 53B74F955C03ADF CRC64;

Query Match 26.6%; Score 55.5; DB 3; Length 1346;
Best Local Similarity 43.8%; Pred. No. 96;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY 3 SYGNSADAPYVGAKIGV---DAKOINGKNT 31
Db 1203 SYGN---OPYKSOIGNIFLNGKVGKST 1230

RESULT 13
005679 ID 005679 PRELIMINARY; PRT; 1346 AA.
AC 005679;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, last annotation update)
DE TY B.
GN YB10822.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S 288C;
RX MEDLINE=96076635; PubMed=7502586;
RA Obermaler B., Gassenhuber J., Piravandi E., Dondely H.;
RT "Sequence analysis of a 78.6 kb segment of the left end of
Saccharomyces cerevisiae chromosome II.";
RL Yeast 11:1103-1112(1995).
DR EMBL: X79489; CAA55998.1; -
DR INTERPRO: IPR001584; -

```
DR INTERPRO: IPR001969; -.
DR PFAM: PF00665; rve; 1.
DR PROSITE: PS00141; ASP_PROTEASE; 1.
DR Hydrolyase; Aspartyl protease.
SQ SEQUENCE 1346 AA; 153931 MW; FF7F64800F9CAFE1 CRC64;

Query Match
Best Local Similarity 43.8%; Score 55.5; DB 3; Length 1346;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

QY 3 SYGNSADAOPYGAKIGOV---DAKOINGKNT 31
Db 1203 SYGN---QPYKSOIGNIFLNGKIVIGKST 1230

RESULT 14
005369 PRELIMINARY: PRT: 1346 AA.
AC 005369;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE TRANSPOSOM TY1-17 154.0 KDA HYPOTHETICAL PROTEIN.
GN TYB.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxId=4932;
RN [1]

SEQUENCE FROM N.A.
RC STRAIN-AB972;
RA Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,
RA Sasanuma S., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,
RA Yamazaki M., Tashiro H., Eki T.;
RT "Analysis of the nucleotide sequence of chromosome VI from
RT Saccharomyces cerevisiae";
RL Nat. Genet. 10:261-268(1995).
DR EMBL: D50617; BAA09237.1; -.
DR INTERPRO: IPR00194; -.
DR PROSITE: PS001584; -.
DR INTERPRO: IPR001969; -.
DR PFAM: PF00665; rve; 1.
DR PROSITE: PS00141; ASP_PROTEASE; 1.
DR PROSITE: PS00152; ATPASE_ALPHA_BETA; UNKNOWN_1.
DR Hydrolyase; Aspartyl protease.
SQ SEQUENCE 1346 AA; 153890 MW; 2A71327338367C21 CRC64;

Query Match
Best Local Similarity 43.8%; Score 55.5; DB 3; Length 1346;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

QY 3 SYGNSADAOPYGAKIGOV---DAKOINGKNT 31
Db 1203 SYGN---QPYKSOIGNIFLNGKIVIGKST 1230

RESULT 15
003494 PRELIMINARY: PRT: 1347 AA.
AC 003494;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE HYPOTHETICAL 153.9 KDA PROTEIN.
GN TYB.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxId=4932;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN-AB972;
RA Oliver K., Shore L., Harris D.;
RL Submitted (DEC-1995) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-AB972;
RA Barrett B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (DEC-1995) to the EMBL/Genbank/DBJ databases.
DR EMBL: Z68194; CAA92351.1; -.
DR INTERPRO: IPR00194; -.
DR INTERPRO: IPR001584; -.
DR INTERPRO: IPR001969; -.
DR PFAM: PF00665; rve; 1.
DR PROSITE: PS00141; ASP_PROTEASE; 1.
DR PROSITE: PS00152; ATPASE_ALPHA_BETA; UNKNOWN_1.
DR Hypothetical protein; Hydrolyase; Aspartyl protease.
SQ SEQUENCE 1347 AA; 153944 MW; 794BCFEB70B06FBB CRC64;
```

```
Query Match
Best Local Similarity 43.8%; Score 55.5; DB 3; Length 1347;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

QY 3 SYGNSADAOPYGAKIGOV---DAKOINGKNT 31
Db 1204 SYGN---QPYKSOIGNIFLNGKIVIGKST 1231
```

Search completed: March 6, 2001, 14:38:26
Job time: 129 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 5, 2001, 01:22:14 ; Search time 1331.97 Seconds
(without alignments)
5588.169 Million cell updates/sec

Title: US-09-164-714-6
Perfect score: 543
Sequence: 1 atgaaacttaaaacact.....gcgcatttgcttta 543

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 1118133 seqs, 6853842396 residues

Total number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_bal:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_ph:*
6: gb_pl1:*
7: gb_pl2:*
8: gb_pl1:*
9: gb_pl2:*
10: gb_pr3:*
11: gb_ro:*
12: gb_sy:*
13: gb_un:*
14: em_fun:*
15: em_hum1:*
16: em_hum2:*
17: em_in:*
18: em_om:*
19: em_or:*
20: em_ov:*
21: em_pat:*
22: em_ph:*
23: em_pl:*
24: em_ro:*
25: em_sts:*
26: em_sy:*
27: em_un:*
28: em_vl:*
29: gb_hlg1:*
30: gb_hlg2:*
31: gb_in1:*
32: gb_in2:*
33: em_ba1:*
34: em_ba2:*
35: em_hum3:*
36: em_hum4:*
37: gb_pr4:*
38: gb_hcg3:*
39: gb_hcg4:*
40: gb_hcg5:*
41: gb_hcg6:*
42: gb_hcg7:*
43: em_hcg1:*

44: em_hcg2:*
45: em_hcg3:*
46: em_hum5:*
47: gb_pl3:*
48: gb_pr5:*
49: gb_hcg8:*
50: gb_hcg9:*
51: gb_hcg10:*
52: gb_hcg11:*
53: gb_hcg12:*
54: gb_hcg13:*
55: gb_hcg14:*
56: gb_in3:*
57: gb_hcg15:*
58: gb_hcg16:*
59: gb_hcg17:*
60: em_hcg4:*
61: em_hcg5:*
62: em_hcg6:*
63: em_hcg7:*
64: em_hum6:*
65: gb_hcg18:*
66: gb_hcg19:*
67: gb_hcg20:*
68: gb_hcg21:*
69: gb_hcg22:*
70: gb_hcg23:*
71: gb_vl1:*
72: gb_vl2:*
73: gb_ba3:*
74: em_hcg8:*
75: em_hcg9:*
76: em_hcg10:*
77: gb_pr6:*
78: gb_pr7:*
79: gb_sts1:*
80: gb_sts2:*
81: gb_pat1:*
82: gb_pat2:*
83: em_hcg0:*
84: gb_hcg24:*
85: gb_pr8:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	8.3	1164	56	TTTAP1PA
2	45	8.3	1575	56	TTTAP1IN
3	44.4	8.2	119118	78	HSRRYR7CC1
4	43.6	8.0	823	7	CAMODUT1
5	43.6	8.0	189286	9	AC008125
6	43.6	8.0	190149	30	AC009770
7	43.6	8.0	213140	54	AC040911
8	42.4	7.8	134741	53	AC026243
9	42.2	7.8	310779	29	AC005140
10	41.8	7.7	121703	59	AC073578
11	41.6	7.7	1302	10	AF090946
12	40.8	7.5	1277	47	SCU20616
13	40.4	7.4	51768	59	AC074318
14	40.4	7.4	58451	1	AC025591
15	40.2	7.4	10976	1	AE001127
16	39.4	7.3	846	81	128501
17	39.4	7.3	873	1	AF077664
18	39.4	7.3	891	81	128499
19	39.4	7.3	934	2	BBOSPBB31
20	39.4	7.3	934	2	BBOSPBBP
21	39.4	7.3	934	2	BBOSPBBEV
					X76126 T.thermophil
					X76125 T.thermophil
					AL049760 Human DNA
					X77925 C.albicans
					AC008125 Homo sapi
					AC009770 Homo sapi
					AC040911 Homo sapi
					AC026243 Homo sapi
					AC005140 Plasmodiu
					AC073578 Homo sapi
					AF090946 Homo sapi
					U20616 Saccharomyc
					AC074318 Staphyloc
					AC025591 Staphyloc
					AE001127 Borrelia
					128501 Sequence 21
					AF077664 Borrelia
					128499 Sequence 19
					X74808 B.burgdorfe
					X74809 B.burgdorfe
					X74810 B.burgdorfe

22	39.4	7.3	1653	2	BOROSPAC	L23137	Borrelia	bu
23	39.4	7.3	1653	2	BOROSPAD	L23138	Borrelia	bu
24	39.4	7.3	1653	2	BOROSPAE	L23139	Borrelia	bu
25	39.4	7.3	1653	2	BOROSPAH	L23141	Borrelia	bu
26	39.4	7.3	1653	2	BOROSPAI	L23142	Borrelia	bu
27	39.4	7.3	1653	2	BOROSPAJ	L23143	Borrelia	bu
28	39.4	7.3	1653	2	BOROSPAK	L23144	Borrelia	bu
29	39.4	7.3	1915	2	BROSPAB	X14407	Borrelia	bu
30	39.4	7.3	1915	81	AR04009	AR04009	B. burgdorferi	
31	39.4	7.3	1916	2	BROSPAB	L19701	Borrelia	bu
32	39.4	7.3	1959	81	AR016731	AR016731	Sequence	
33	39.4	7.3	5952	81	AR064321	AR064321	Sequence	
34	39.4	7.3	53561	1	AE000790	AE000790	Borrelia	bu
35	39.2	7.2	195697	38	AC010707	AC010707	Drosophila	
36	39	7.2	136410	50	AC023308	AC023308	Homo sapi	
37	39	7.2	152355	9	AC005668	AC005668	Homo sapi	
38	38.8	7.1	198570	77	HS343C1	AL008720	Human DNA	
39	38.6	7.1	86064	6	AB013395	AB013395	Arabidops	
40	38.6	7.1	192518	54	AC027745	AC027745	Homo sapi	
41	38.6	7.1	197715	52	AC025493	AC025493	Homo sapi	
42	38.2	7.0	566	2	RCU45244	U45244	Rickettsia	
43	38.2	7.0	566	2	RCU46918	U46918	Rickettsia	
44	37.8	7.0	873	1	AF077663	AF077663	Borrelia	bu
45	37.8	7.0	891	2	BOROSPBVR	L31399	Borrelia	bu

ALIGNMENTS

RESULT 1	TTTAP1PA	1164 bp	mrna	INV	02-SEP-1994
LOCUS	TTTAP1PA	T.thermophila TAP1 gene	polya signal with multiple stop codons.		
DEFINITION	X76126	GI:426480			
ACCESSION	X76126.1	GI:426480			
VERSION	1				
KEYWORDS	tap1 gene.				
SOURCE	Tetrahymena thermophila.				
ORGANISM	Macronuclear Tetrahymena thermophila				
REFERENCE	Eukaryota: Alveolata: Ciliophora: Oligohymenophorea;				
AUTHORS	Hymenostomatida; Tetrahymenina; Tetrahymena.				
TITLE	1 (bases 1 to 1164)				
JOURNAL	Pearlman, R.E.				
REFERENCE	Direct Submission				
AUTHORS	Submitted (05-NOV-1993) R.E. Pearlman, York University, Dept. of				
TITLE	Biology, Farquharson Bldg., 4700 Keele St., Downsview, Ontario M3J				
JOURNAL	1P3, CANADA				
REFERENCE	2 (bases 1 to 1164)				
AUTHORS	Heinonen, T.Y. and Pearlman, R.E.				
TITLE	A germ line-specific sequence element in an intron in Tetrahymena				
JOURNAL	thermophila				
REFERENCE	J. Biol. Chem. 269 (26), 17428-17433 (1994)				
MEDLINE	94292495				
FEATURES	Location/Qualifiers				
SOURCE	1. 1164				
	/organism="Tetrahymena thermophila"				
	/macronuclear				
	/db_xref="taxon:5911"				
	/chromosome="1"				
	/dev_stage="macronucleus"				
	/clone_lib="plasmid"				
	/note="rearranged"				
	103..1020				
gene	/gene="TAP1"				
	103..105				
misc_feature	/gene="TAP1"				
	961..963				
misc_feature	/note="stop codon TAA-Q"				
	967..969				
misc_feature	/gene="TAP1"				
	/note="stop codon TAA-Q"				
	997..999				
misc_feature					

RESULT 2	TTTAP1IN	1575 bp	DNA	INV	29-SEP-1998
LOCUS	TTTAP1IN	T.thermophila TAP1 gene.			
DEFINITION	X76125	GI:426479			
ACCESSION	X76125.1	GI:426479			
VERSION	1				
KEYWORDS	tap1 gene.				
SOURCE	Tetrahymena thermophila.				
ORGANISM	Macronuclear Tetrahymena thermophila				
REFERENCE	Eukaryota: Alveolata: Ciliophora: Oligohymenophorea;				
AUTHORS	Hymenostomatida; Tetrahymenina; Tetrahymena.				
TITLE	1 (bases 1 to 1575)				
JOURNAL	Pearlman, R.E.				
REFERENCE	Direct Submission				
AUTHORS	Submitted (05-NOV-1993) R.E. Pearlman, York University, Dept. of				
TITLE	Biology, Farquharson Bldg., 4700 Keele St., Downsview, Ontario M3J				
JOURNAL	1P3, CANADA				
REFERENCE	2 (bases 1 to 1575)				
AUTHORS	Heinonen, T.Y. and Pearlman, R.E.				
TITLE	A germ line-specific sequence element in an intron in Tetrahymena				
JOURNAL	thermophila				
REFERENCE	J. Biol. Chem. 269 (26), 17428-17433 (1994)				
MEDLINE	94292495				
FEATURES	Location/Qualifiers				
SOURCE	1. 1575				
	/organism="Tetrahymena thermophila"				
	/macronuclear				
	/db_xref="taxon:5911"				
	/chromosome="1"				
	/dev_stage="macronucleus"				
	/clone_lib="plasmid"				
	/note="rearranged"				
	join<16..205,383..664,899..>1431)				
mrna	/gene="TAP1"				
	16..1555				
gene	/gene="TAP1"				
	<16..205				
exon	/gene="TAP1"				
	/number=1				
CDS	join(16..205,383..664,899..1431)				
	/gene="TAP1"				

Query Match

Best local similarity 8.3%; Score 45; DB 56; Length 1164;

Matches 108; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

Qy	83	atctcgtatgctcaccatagctgtggtgcacaaatgltcaagtacgacgaagcaaa	142	/gene="TAP1"
Db	653	ATGCTGATGATGCTGTAAGACGGTGATGATGCCGAAGATGCTGATTAACGAAGATGCCG	712	/note="stop codon TAA-Q"
Qy	143	tcaacggtagaacacccgtatggtatctatgcaggtataacttgaccacaaatttg	202	/gene="TAP1"
Db	713	AAGATGATGATGATGCTGAAGATGATGATGCCGAAGATGCTGATTAACGAAGATG	772	/note="stop codon TGA"
Qy	203	gcgtagaaccggaattggtgttcaagacgccaagaatttaatgacggcgtgagtcctg	262	/note="addition"
Db	773	GGCATGATGCCGAAGATGCTGATGATGCCGAAGATGCTGATGCCGAAGATGCTGATG	832	
Qy	263	taaaagtgatgtgaagctcttcttggtcctatg	295	
Db	833	ATAACGAAGATGCCGAAGATGCTGATGATGCTG	865	


```

repeat_region      6501..6797
                    /note="AluY repeat: matches 1..297 of consensus"
repeat_region      7334..7400
                    /note="MER21B repeat: matches 706..789 of consensus"
repeat_region      7779..7894
                    /note="MER2 repeat: matches 18..132 of consensus"
repeat_region      7895..7975
                    /note="AluY repeat: matches 2..82 of consensus"
repeat_region      8003..8080
                    /note="MER2 repeat: matches 269..343 of consensus"
repeat_region      8105..8229
                    /note="AluSg/x repeat: matches 9..133 of consensus"
repeat_region      8257..8415
                    /note="AluYo/FRAM repeat: matches 149..304 of consensus"
repeat_region      8630..8919
                    /note="AluYo repeat: matches 13..300 of consensus"
repeat_region      9266..9580
                    /note="AluSp repeat: matches 5..313 of consensus"
repeat_region      9714..9845
                    /note="MIR repeat: matches 105..262 of consensus"
repeat_region      9973..10170
                    /note="MIR repeat: matches 34..240 of consensus"
repeat_region      10270..10480
                    /note="L1MA10 repeat: matches 6135..6322 of consensus"
repeat_region      10481..10735
                    /note="AluSg repeat: matches 57..311 of consensus"
repeat_region      10736..10821
                    /note="L1MA10 repeat: matches 6045..6135 of consensus"
repeat_region      10871..11171
                    /note="AluYo repeat: matches 1..303 of consensus"
repeat_region      11190..11487
                    /note="AluX repeat: matches 1..298 of consensus"
repeat_region      11610..11726
                    /note="MIR repeat: matches 13..134 of consensus"
repeat_region      12322..12636
                    /note="AluX repeat: matches 1..312 of consensus"
repeat_region      13016..13133
                    /note="L2 repeat: matches 2616..2748 of consensus"
repeat_region      13419..13704
                    /note="L13 copies 2 mer 99 59 conserved"
repeat_region      13432..13769
                    /note="L13 copies 26 mer 57 conserved"
repeat_region      13878..14208
                    /note="MLTIC repeat: matches 6..368 of consensus"
repeat_region      14224..14886
                    /note="L1PA10 repeat: matches 5476..6165 of consensus"
repeat_region      14893..15204
                    /note="AluSp repeat: matches 1..312 of consensus"
repeat_region      15226..15335
                    /note="L1P3 repeat: matches 5356..5465 of consensus"
repeat_region      15336..15446
                    /note="MLTIC repeat: matches 355..464 of consensus"
repeat_region      15484..15777
                    /note="AluSg repeat: matches 1..295 of consensus"
misc_feature      complement(15768..16376)
                    /note="match: GSS: Em:AQ476535"
repeat_region      16387..16541
                    /note="MIR repeat: matches 73..226 of consensus"
repeat_region      17071..17122
                    /note="L26 copies 2 mer aa 73 conserved"
repeat_region      17232..17308
                    /note="L2 repeat: matches 2652..2738 of consensus"
repeat_region      17452..17487
                    /note="L18 copies 2 mer tg 91 conserved"
repeat_region      17599..17908
                    /note="AluY repeat: matches 1..310 of consensus"
repeat_region      18508..18616
                    /note="L1PA6 repeat: matches 6035..6143 of consensus"
repeat_region      18668..18865
                    /note="AluX repeat: matches 1..298 of consensus"
repeat_region      19069..19100
                    /note="MLT1-INTERNAL repeat: matches 692..723 of consensus"

```

```

repeat_region      19494..19849
                    /note="MLT1-INTERNAL repeat: matches 927..1296 of consensus"
repeat_region      19869..20149
                    /note="AluYo repeat: matches 18..288 of consensus"
repeat_region      20218..20385
                    /note="MLT1B repeat: matches 1..180 of consensus"
repeat_region      20386..20684
                    /note="AluSp repeat: matches 1..303 of consensus"
repeat_region      20685..20736
                    /note="MLT1B repeat: matches 180..234 of consensus"
repeat_region      20737..21016
                    /note="AluSg repeat: matches 1..280 of consensus"
repeat_region      21017..21215
                    /note="MLT1B repeat: matches 234..390 of consensus"
repeat_region      21217..21347
                    /note="MIR repeat: matches 12..144 of consensus"
repeat_region      21582..21731
                    /note="L2 repeat: matches 2311..2657 of consensus"
repeat_region      22386..22415
                    /note="L15 copies 2 mer ac 93 conserved"
repeat_region      23579..23881
                    /note="AluX repeat: matches 1..310 of consensus"
misc_feature      complement(23985..24510)
                    /note="match: GSS: Em:AQ529159"
misc_feature      complement(24086..24604)
                    /note="match: GSS: Em:AQ569063"
repeat_region      24115..24201
                    /note="MLT1D repeat: matches 6..94 of consensus"
repeat_region      24201..24304
                    /note="MLT1B repeat: matches 362..466 of consensus"
misc_feature      complement(24215..24582)
                    /note="match: GSS: Em:AQ375812"
misc_feature      24487..24919
                    /note="match: GSS: Em:AQ206604"

Query Match      8.2%; Score 44.4; DB 78; Length 119118;
Best Local Similarity 49.2%; Pred.No. 0.42;
Matches 117; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

QY      74 gctatggcaattctgctgtagtgcacccatagttggtgcacaaattgtaagtacgacg 133
         ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
DB 73202 GTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 73143

QY      134 ccaagcaatcaacgyltaagaacacccgtatgtaattatgtaacaggtltaacttgacc 193
         ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 73142 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 73083

QY      194 aaaaatttgctagaacccgaattgtgtgttcgagcccaagaattaatgcagcgcg 253
         || || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
DB 73082 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 73023

QY      254 tgaagctgtlaaaagtgatgaaagcttcttggttgtaattggtacacatacgclatba 311
         || || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
DB 73022 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 72565

RESULT 4
CAMODUT1      823 bp      DNA      PLN      03-AUG-1995
LOCUS      C.albicans (WO-1) DUT1 gene.
DEFINITION      X77925
ACCESSION      X77925.1 GI:457720
VERSION      DUT1 gene; dUTPase.
KEYWORDS      Candida albicans.
SOURCE      Candida albicans.
ORGANISM      Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
              Candidaceae; Candida.
REFERENCE      1 (bases 1 to 823)
AUTHORS      McIntosh,E.M., Looser,J., Haynes,R.H. and Pearlman,R.E.
TITLE      Mui site-dependent transcriptional regulation of the Candida
              albicans dUTPase gene
JOURNAL      Curr. Genet. 26 (5-6), 415-421 (1994)

```

MEDLINE	95197975
REMARK	Erratum: [[published erratum appears in Curr Genet 1995 Apr;27(5):491]]
REFERENCE	2 (bases 1 to 823)
AUTHORS	Pearlman,R.E.
TITLE	Direct Submission
JOURNAL	Submitted (28-FEB-1994) R.E. Pearlman, York University

BASE COUNT	290 a	138 c	146 g	249 t
ORIGIN				

AC008125
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS

TITLE	.Direct Submission
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 189286)
AUTHORS	Worley,K.C.
TITLE	Direct Submission
JOURNAL	Submitted (24-JUL-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE	3 (bases 1 to 189286)
AUTHORS	Worley,K.C.
TITLE	Direct Submission
JOURNAL	Submitted (25-SEP-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT	On Sep 25, 1999 this sequence version replaced gi:5851701.

ORGANISM Homo sapiens
 Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 213140)
 AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 TITLE Homo sapiens chromosome 1, clone RP11-217N8
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 213140)
 AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
 Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G.,
 Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
 Collymore, A., Cooke, P., Dearrellano, K., Dewar, K., Diaz, J.S.,
 Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
 Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
 Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
 Klein, J., Lacroque, K., Lamazares, R., Landers, T., Lehoczy, J.,
 Levine, R., Lieu, C., Liu, G., Locke, K., MacDonald, P., Margulis, N.,
 McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheters, R.,
 Meldrum, J., Menes, L., Milnova, T., Miranda, C., Mlenga, V., Morrow, J.,
 Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
 Pisanil, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
 Stange-Thomann, N., Stojanovic, N., Sudramanian, A., Talamas, J.,
 Teefaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
 Young, C., Zainoun, J., Zimmer, A. and Zody, M.
 Submitted (11-APR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

 Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIMR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu

 Project Information
 Center project name: L7715
 Center clone name: 217.N.8

 Summary Statistics
 Sequencing vector: M13; M77815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 182556 bases at least Q40
 Consensus quality: 198779 bases at least Q30
 Consensus quality: 204626 bases at least Q20
 Insert size: 18200; agarose-fp
 Insert size: 207940; sum-of-ctrls
 Quality coverage: 3.9 in Q20 bases; agarose-fp
 Quality coverage: 3.4 in Q20 bases; sum-of-ctrls

 NOTE: This is a 'working draft' sequence. It currently
 * consists of 53 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1 1214: contig of 1214 bp in length
 * 1215 1314: gap of 100 bp
 * 1315 2372: contig of 1058 bp in length
 * 2373 2472: gap of 100 bp
 * 2473 3495: contig of 1023 bp in length
 * 3496 3593: gap of 100 bp
 * 3596 4625: contig of 1030 bp in length
 * 4626 4725: gap of 100 bp
 * 4726 5839: contig of 1114 bp in length
 * 5840 5939: gap of 100 bp

* 5940 7014: contig of 1075 bp in length
 * 7015 7114: gap of 100 bp
 * 7115 8180: contig of 1066 bp in length
 * 8181 8280: gap of 100 bp
 * 8281 9604: contig of 1324 bp in length
 * 9605 9704: gap of 100 bp
 * 9705 11339: contig of 1635 bp in length
 * 11340 11439: gap of 100 bp
 * 11440 12708: contig of 1269 bp in length
 * 12709 12808: gap of 100 bp
 * 12809 14202: contig of 1304 bp in length
 * 14203 14302: gap of 100 bp
 * 14303 15487: contig of 1185 bp in length
 * 15488 15587: gap of 100 bp
 * 15588 16909: contig of 1322 bp in length
 * 16910 17009: gap of 100 bp
 * 17010 18251: contig of 1242 bp in length
 * 18252 18351: gap of 100 bp
 * 18352 19767: contig of 1416 bp in length
 * 19768 19867: gap of 100 bp
 * 19868 21172: contig of 1305 bp in length
 * 21173 21272: gap of 100 bp
 * 21273 23727: contig of 2455 bp in length
 * 23728 23827: gap of 100 bp
 * 23828 24931: contig of 1104 bp in length
 * 24932 25031: gap of 100 bp
 * 25032 26475: contig of 1444 bp in length
 * 26476 26575: gap of 100 bp
 * 26576 28259: contig of 1684 bp in length
 * 28260 28359: gap of 100 bp
 * 28360 31078: contig of 2719 bp in length
 * 31079 31178: gap of 100 bp
 * 31179 32716: contig of 1538 bp in length
 * 32717 32816: gap of 100 bp
 * 32817 33881: contig of 1065 bp in length
 * 33882 33981: gap of 100 bp
 * 33982 36218: contig of 2237 bp in length
 * 36219 36318: gap of 100 bp
 * 36319 37461: contig of 1143 bp in length
 * 37462 37561: gap of 100 bp
 * 37562 40250: contig of 2689 bp in length
 * 40251 40350: gap of 100 bp
 * 40351 41626: contig of 1276 bp in length
 * 41627 41726: gap of 100 bp
 * 41727 44548: contig of 2822 bp in length
 * 44549 44648: gap of 100 bp
 * 44649 46762: contig of 2114 bp in length
 * 46763 46862: gap of 100 bp
 * 46863 49156: contig of 2294 bp in length
 * 49157 49256: gap of 100 bp
 * 49257 51320: contig of 2064 bp in length
 * 51321 51420: gap of 100 bp
 * 51421 53478: contig of 2058 bp in length
 * 53479 53578: gap of 100 bp
 * 53579 56891: contig of 3313 bp in length
 * 56892 56991: gap of 100 bp
 * 56992 59163: contig of 2172 bp in length
 * 59164 59263: gap of 100 bp
 * 59264 62101: contig of 2838 bp in length
 * 62102 62201: gap of 100 bp
 * 62202 65836: contig of 3635 bp in length
 * 65837 65936: gap of 100 bp
 * 65937 69543: contig of 3607 bp in length
 * 69544 69643: gap of 100 bp
 * 69644 74765: contig of 5122 bp in length
 * 74766 74865: gap of 100 bp
 * 74866 78671: contig of 3806 bp in length
 * 78672 78771: gap of 100 bp
 * 78772 83032: contig of 4261 bp in length
 * 83033 83132: gap of 100 bp
 * 83133 90972: contig of 7840 bp in length
 * 90973 91072: gap of 100 bp
 * 91073 97926: contig of 6854 bp in length

* 4369 4468: gap of 100 bp
* 4469 5268: contig of 800 bp in length
* 5269 5368: gap of 100 bp
* 5369 6147: contig of 779 bp in length
* 6148 6247: gap of 100 bp
* 6248 7036: contig of 789 bp in length
* 7037 7136: gap of 100 bp
* 7137 7922: contig of 786 bp in length
* 7923 8022: gap of 100 bp
* 8023 8819: contig of 797 bp in length
* 8820 8919: gap of 100 bp
* 8920 9722: contig of 803 bp in length
* 9723 9822: gap of 100 bp
* 9823 10604: contig of 782 bp in length
* 10605 10704: gap of 100 bp
* 10705 11466: contig of 762 bp in length
* 11467 11566: gap of 100 bp
* 11567 12363: contig of 797 bp in length
* 12364 12463: gap of 100 bp
* 12464 13250: contig of 787 bp in length
* 13251 13350: gap of 100 bp
* 13351 14130: contig of 780 bp in length
* 14131 14230: gap of 100 bp
* 14231 15030: contig of 800 bp in length
* 15031 15130: gap of 100 bp
* 15131 15928: contig of 798 bp in length
* 15929 16028: gap of 100 bp
* 16029 16809: contig of 781 bp in length
* 16810 16909: gap of 100 bp
* 16910 17701: contig of 792 bp in length
* 17702 17801: gap of 100 bp
* 17802 18586: contig of 785 bp in length
* 18587 18686: gap of 100 bp
* 18687 19502: contig of 816 bp in length
* 19503 19602: gap of 100 bp
* 19603 20406: contig of 804 bp in length
* 20407 20506: gap of 100 bp
* 20507 21275: contig of 769 bp in length
* 21276 21375: gap of 100 bp
* 21376 22168: contig of 793 bp in length
* 22169 22268: gap of 100 bp
* 22269 23073: contig of 805 bp in length
* 23074 23173: gap of 100 bp
* 23174 23974: contig of 801 bp in length
* 23975 24074: gap of 100 bp
* 24075 24839: contig of 765 bp in length
* 24840 24939: gap of 100 bp
* 24940 25705: contig of 766 bp in length
* 25706 25805: gap of 100 bp
* 25806 26600: contig of 795 bp in length
* 26601 26700: gap of 100 bp
* 26701 27488: contig of 788 bp in length
* 27489 27588: gap of 100 bp
* 27589 28373: contig of 785 bp in length
* 28374 28473: gap of 100 bp
* 28474 29278: contig of 805 bp in length
* 29279 29378: gap of 100 bp
* 29379 30187: contig of 809 bp in length
* 30188 30287: gap of 100 bp
* 30288 31062: contig of 775 bp in length
* 31063 31162: gap of 100 bp
* 31163 31949: contig of 787 bp in length
* 31950 32049: gap of 100 bp
* 32050 32841: contig of 792 bp in length
* 32842 32941: gap of 100 bp
* 32942 33725: contig of 784 bp in length
* 33726 33825: gap of 100 bp
* 33826 34633: contig of 808 bp in length
* 34634 34733: gap of 100 bp
* 34734 35515: contig of 782 bp in length
* 35516 35615: gap of 100 bp
* 35616 36404: contig of 789 bp in length
* 36405 36504: gap of 100 bp

* 35505 37288: contig of 784 bp in length
* 37289 37388: gap of 100 bp
* 37389 38185: contig of 797 bp in length
* 38186 38285: gap of 100 bp
* 38286 39090: contig of 805 bp in length
* 39091 39190: gap of 100 bp
* 39191 39985: contig of 795 bp in length
* 39986 40085: gap of 100 bp
* 40086 40867: contig of 782 bp in length
* 40868 40967: gap of 100 bp
* 40968 41747: contig of 780 bp in length
* 41748 41847: gap of 100 bp
* 41848 42615: contig of 768 bp in length
* 42616 42715: gap of 100 bp
* 42716 43495: contig of 780 bp in length
* 43496 43595: gap of 100 bp
* 43596 44384: contig of 789 bp in length
* 44385 44484: gap of 100 bp
* 44485 45281: contig of 797 bp in length
* 45282 45381: gap of 100 bp
* 45382 46189: contig of 808 bp in length
* 46190 46289: gap of 100 bp
* 46290 47085: contig of 796 bp in length
* 47086 47185: gap of 100 bp
* 47186 47956: contig of 771 bp in length
* 47957 48056: gap of 100 bp
* 48057 48846: contig of 790 bp in length
* 48847 48946: gap of 100 bp
* 48947 49745: contig of 799 bp in length
* 49746 49845: gap of 100 bp
* 49846 50629: contig of 784 bp in length
* 50630 50729: gap of 100 bp
* 50730 51519: contig of 790 bp in length
* 51520 51619: gap of 100 bp
* 51620 52398: contig of 779 bp in length
* 52399 52498: gap of 100 bp
* 52499 53250: contig of 752 bp in length
* 53251 53350: gap of 100 bp
* 53351 54154: contig of 804 bp in length
* 54155 54254: gap of 100 bp
* 54255 55065: contig of 811 bp in length
* 55066 55165: gap of 100 bp
* 55166 55949: contig of 784 bp in length
* 55950 56049: gap of 100 bp
* 56050 56854: contig of 805 bp in length
* 56855 56954: gap of 100 bp
* 56955 57752: contig of 798 bp in length
* 57753 57852: gap of 100 bp
* 57853 58691: contig of 839 bp in length
* 58692 58791: gap of 100 bp
* 58792 59590: contig of 799 bp in length
* 59591 59690: gap of 100 bp
* 59691 60510: contig of 820 bp in length
* 60511 60610: gap of 100 bp
* 60611 61410: contig of 800 bp in length
* 61411 61510: gap of 100 bp

Query Match 7.8% Score 42.4; DB 53; Length 134741;
Best Local Similarity 47.7%; Pred. No. 1.4;
Matches 124; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

QY 56 gfgctaaagctccatcagctatggaattcgtcgtatgctcaaccctatgtgtgca 115
DB 89095 GTGATGATGCTGATGCTGATGATGCTGATGATGATGATGCTGATGATG 89036
QY 116 aaattgtaagtaagcgaacgaatacaacgtaagaacacggtttgtattatg 175
DB 89035 ATGATGCTTATGCTGCTGATGATGATGCTTATGCTGATGCTGATGATG 88976
QY 176 caggtataactttgaacaaaatlilggcglagaaacccgaatttltgttcaagcca 235
DB 88975 ATGCTTATGCACTGACGATGCTGATGCTGATGCTGATGCTGATGCTG 88916


```

/product="G4p2"
/protein_id="AAA70169.1"
/db_xref="GI:710338"
/translation="MSNPEDLLGNVEDADVVLPKEILVKSNTSKKADVPSPADP
SKARNRPPSGNMGARIDKTRGRNRNRKSDJTSATKKSNTRRAVDHSHSTGKTDY
KKKVQMGDDKKEISAEKEADADAEIAEDAEEDGKPKTALSLQDILNOOAN
NOENFVPEAKREVELDARGLETAKEAYVATVKNVKSQKLTREKLEEDAEVESENT
RNFQDRNRNSNNFNRRGRGARGKRNNTANATNSANTVQRNRNIDVSNLPSLA"
BASE COUNT      415 a      262 c      250 g      350 t
ORIGIN

```

```

Query Match      7.5%; Score 40.8; DB 47; Length 1277;
Best Local Similarity 44.7%; Pred. No. 3.3;
Matches 159; Conservative 0; Mismatches 197; Indels 0; Gaps 0;

```

```

QY  91 gatgctcaaccctatgttggtgccaataatggtccaagttagccgaacaaatcaacggt 150
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  597 GAAGCTCAAGCCGATCTGCTGTAATTTGCTGAAGACCTCAGAAAGCTGAAGAGCT 656
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  151 aagaacaccgcttatgtattatgcaagttataactcttgacaaaatttgcgtagaa 210
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  657 GGAAGCCAAAGACCCCTCAATTTCTTTGCAAGACTACTTGAAACCAACAAGCTAACAC 716
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  211 cccgaatttggttgcagacgccaagaattaatgcaagcgtagtccgtgaaaggt 270
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  717 CAGTTCACAGAGTCCCGACAGAACTAAGAGAGTGAATTAGACCGTGAAGAAATTTGAAC 776
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  271 gatgtgaagtcttctgtgtctatgacacatcgctataacttcatacccaattt 330
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  777 GCTGAAGAGAGAGCTTACGTTCCAGCAACCAAGCTGACAGCTCAATTAAGCAATTG 836
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  331 tatgccaagggcaaatatgagcattgctgaagtaagtagatgttaccgacgtaagca 390
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  837 AAGACCAAGAGAGTACTTGCAATTTGATGCCACTTTTGTTCATTAACCTAGAAAAG 896
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  391 actacatactcaacaaagcgacaaacagccttagcagcggtgtgtgtgtgtgtgt 446
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  897 TTGCGTGACGAGAAACAACACAGCAGAAACACTTCAACACCGTGTGTGTGTGTGT 952
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 13
AC074318/c
LOCUS
DEFINITION
  Staphylococcus aureus clone sabac-108, WORKING DRAFT SEQUENCE, 1
  ordered pieces.
ACCESSION
  AC074318
VERSION
  AC074318.4 GI:10799432
KEYWORDS
  HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE
  Staphylococcus aureus
  Staphylococcus aureus
  Bacteria; Firmicutes; Bacillus/Clostridium group;
  Bacillus/staphylococcus group; Staphylococcus.
  1 (bases 1 to 51788)
REFERENCE
  Tian, R., Lin, S., Jia, H., Qian, Y., Iondola, J. and Roe, B.A.
  Staphylococcus aureus BAC Clone sabac-108
  Unpublished
  2 (bases 1 to 51788)
AUTHORS
  Tian, R., Lin, S., Jia, H., Qian, Y., Iondola, J. and Roe, B.A.
  Direct Submission
  Submitted (26-JUL-2000) Department Of Chemistry And Biochemistry,
  The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
  OK 73019, USA
COMMENT
  On Oct 12, 2000 this sequence version replaced gi:10765076.
  * NOTE: This is a 'working draft' sequence. It currently
  * consists of 1 contigs. Gaps between the contigs
  * are represented as runs of N. The order of the pieces
  * is believed to be correct as given, however the sizes
  * of the gaps between them are based on estimates that have
  * provided by the submittor.
  * This sequence will be replaced
  * by the finished sequence as soon as it is available and
  * the accession number will be preserved.

```

```

FEATURES
  *
  1 51788: contig of 51788 bp in length.
  Location/Qualifiers
  source
    /organism="Staphylococcus aureus"
    /db_xref="taxon:1280"
    /clone="sabac-108"
    /clone_1fb="OU staph library"
    /clone_1fb="OU staph library"
BASE COUNT      16458 a      9849 c      7959 g      17522 t
ORIGIN

```

```

Query Match      7.4%; Score 40.4; DB 59; Length 51788;
Best Local Similarity 51.7%; Pred. No. 4.4;
Matches 92; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

```

```

QY  91 gatgctcaaccctatgttggtgccaataatggtccaagttagccgaacaaatcaacggt 150
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  3119 GATGCTCAGTATTCTATGATGGAAGACATGGTACCTTTATACCAACAGCTTGAAATTG 3060
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  151 aagaacaccgcttatgtattatgcaagttataactcttgacaaaatttgcgtagaa 210
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  3059 GAAACCAAGCGCGTCGACTTTTATGAAGGTATGAAGTGAAGTGAAGTGAAGTGAAG 3000
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  211 cccgaatttggttgcagacgccaagaattaatgcaagcgtagtccgtgaaaggt 268
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  2999 AAACCTTGATTTGATGAGTGTGAATGATGATGATGATGATGATGATGATGATGATG 2942
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 14
AC025591
LOCUS
DEFINITION
  Staphylococcus aureus clone sabac-134, complete sequence.
ACCESSION
  AC025591
VERSION
  AC025591.7 GI:8570488
KEYWORDS
  HTG.
SOURCE
  Staphylococcus aureus
  Bacteria; Firmicutes; Bacillus/Clostridium group;
  Bacillus/staphylococcus group; Staphylococcus.
  1 (bases 1 to 58431)
REFERENCE
  Loh, P., Qi, S., Ray, L., Ford, B., Iondola, J. and Roe, B.A.
  Staphylococcus aureus BAC sabac-134
  Unpublished
  2 (bases 1 to 58431)
AUTHORS
  Loh, P., Qi, S., Ray, L., Ford, B., Iondola, J. and Roe, B.A.
  Direct Submission
  Submitted (12-MAR-2000) Department Of Chemistry And Biochemistry,
  The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
  OK 73019, USA
COMMENT
  On Jun 17, 2000 this sequence version replaced gi:8567825.
  Location/Qualifiers
  source
    /organism="Staphylococcus aureus"
    /db_xref="taxon:1280"
    /clone="sabac-134"
    /clone_1fb="OU staph library"
BASE COUNT      19747 a      9170 c      10899 g      18615 t
ORIGIN

```

```

Query Match      7.4%; Score 40.4; DB 1; Length 58431;
Best Local Similarity 51.7%; Pred. No. 4.4;
Matches 92; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

```

```

QY  91 gatgctcaaccctatgttggtgccaataatggtccaagttagccgaacaaatcaacggt 150
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  39398 GATGCTCAGTATTCTATGATGGAAGACATGGTACACTTTATACCAACACGTTGAATTG 39457
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

OY	151	aagaacacgcgtctatgcatctatcaggtcttaacttgcacaaatttgcgtaga	210
DB	39458	GAAACAAGAGCGCTGACCTTTTATGAGGTGAAGTGGAAGCTGMAAAGACGA	39517
OY	211	cccgacttgcgttcagcagcccaagaatttaagcagcgctgagtcctctaaag	268
DB	39518	AAAGCTTGAGTTGATGACGACTTCTAATATGATCTGACCAAGCTTGTAACCGAACAAG	39575
RESULT	15		
LOCUS	AE001127	10976 bp	DNA
DEFINITION	Borrelia burgdorferi (section 13 of 70) of the complete genome.	15-DEC-1997	
ACCESSION	AE000783		
VERSION	AE001127.1	GI:2688047	
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			

DYSSTIEHSHVAKIRGLKELDELDFPRKATIEVGRGVITNRNADGELPDSIIILNDEAY
 DEYYVSDIILFKPTKYGIELFSSLLTAYKNEISSLLTAVKNEKITIADSDIIVKPEPS
 QKELIYVKSLLIIPESRALYPIEFYKATIFENIASNNVYKDYFVAKIKELKIDIEKTILE
 DEKNITIGIISKIISGILEIKDEVLGRKGITFSKNFVFLERLNFINEKRDLELKA
 EAVKIDESIKAIKLEIETLEKFKKSIDIEEFPGLKVSGLNDPENFLVANNYFEESI
 HDILIKIRIOVENEFYKAFVEFFLEBEYKVLNFSLENDPNSIASIFKILTSFGDPSIF
 TEEGLEEYLSVNSRHYINNECKNKKVYFEDLIYSHRIRCKITQSVOVHKIDIKIYVA
 AKYFNSSNLPITYKIVITLDYSLINAAKRLSIFKRIYFETIGKHSYKALSRDPE
 RAGLKSSSGCLFDSNERVKRQLEKIRKOREIKEDOVNRSPFKKGAAPKOVASVYKQ
 RODEKAMVFRSDRIQCK"
 gene complement(6380. .7393)
 /gene="BB0164"
 complement(6380. .7393)
 /gene="BB0164"
 /note="similar to GB:L77117 SP:O57556 PID:1590872 percent
 identity: 34.47; identified by sequence similarity;
 putative"
 /codon_start=1
 /transl_table=11
 /product="Na+/Ca+ exchange protein, putative"
 /protein_id="AAC66548.1"
 /db_xref="GI:2688049"
 /transl_table="MLFPLVKEIYLQHLIQEYVFGCIFLILYAGNLLKSSVGIATYLL
 KPTPLIICVITIVSETSAPLEFSLVAAFKKNIVASVNGSIINMLLALPLAFGE
 LKIRKDFDRLKSPMEFLFIMLELLISDFGRSSFPVTPYRNSSTIILFESTYLL
 LFYREKRYASLSENGFSGANLNSQNFEPFNLSIPATISWYKFLYESSKILLNCGALY
 ANNVNVSSEKLGIILVAFGTSVPELVYSFLAIIKRESDIARFGNIGTSNFFNGFIILA
 SSSEFPIILDDIYITLDFSIWVFTIYLVLVKRGVFGRIISIFLLLYFLVNLMEF
 NY"
 gene complement(7394. .9238)
 /gene="BB0165"
 complement(7394. .9238)
 /gene="BB0165"
 /note="hypothetical protein; identified by Glimmer:

/codon_start=1
 /transl_table=11
 /product="B. burgdorferi predicted coding region BB0165"
 /protein_id="AAC66547.1"
 /db_xref="GI:2688052"
 /translation="MYMKLFIWVYVLPFSVFKVPSIYSLTDEEFPKKSILFVPHKGF
 LSKNNGITKIVQVGNISRWYVPEFKLPSRTSLIYEDVYSSSELTTSNNLYVSD
 YSKNRKLVIGDIFKNSAIGTSSAFSGOGRYRATIGTAHGIYLVNGALISPEKNIL
 IPIQLGAYGDIISAIKSEKTEKNNLYVSSGVDGFILOSQSGFKRISPEKNIL
 IRIIDLSKSNVEKILVRYTDNHPYSTINQWPIGLSLDDQDFPEKSOHQAKKNG
 FIDVYVILKRAKKAIDVYRIVARCVKDSLKYDNDKHALMNKTKPAHLLIKVDG
 SLKVCYQVEHVVDFSTLPMRQYVNIISAEIGSFQEDIOFDYIFRSDPSVLAISR
 NKMVQYVDADEELSLIMARQELYVIAIDVIGYQWPEPNSIGNQISMLSDVATIS
 PMFYSHITDQLPENFYTKRAYRYRKGSGRALASISLDGVYIRVYQVAFLLKKEVL
 VDDEYILEYLRKQIKGKSESGSLMASNVYTIKSGLKETLDGS"
 complement(9358..10878)
 /gene="BB0165"
 complement(9358..10878)
 /gene="BB0165"
 /note="similar to PID:1651873 percent identity: 38.63;
 identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="4-alpha-glucanotransferase (mal0)"
 /protein_id="AAC66547.1"
 /db_xref="GI:2688048"
 /translation="KKYKKYRNLNLKRSKJILNLSLPSKYGICDLCGAKKFI
 LFASSQIWMQAYSPIDYTRSPYISFAPGNVYITDEALDKFIDLSLILKKE
 RYRSLDKLITSEDFKLKEALNFINNAADEVKSEFKPKSSYWLIDFSSVPAFKY
 EFKSEKANEVYFDGLIKRNDKLFKRLNLSKTEKQVQVLEQYFFESQOALRYVN

Query Match	7.48;	Score 40.2;	DB 1;	Length 10976;
Best Local Similarity	49.38;	Pred. No. 4.8;		
Matches 105; Conservative	0;	Mismatches 108;	Indels 0;	Gaps 0;

```

Db      5750  TTGTGTAATAAAATATTTTATTTATATTCCTCAATAAATAGTTGTAAAGATGTTTTCTT 5691
Oy      223   gttcagacgcccagaagatttaatgcagagtgtagccctgtaabzygbatgtgaagctc 282
Db      5690  AGAATTAAAGCATTTTAAAGCTTATTGATTATGACAAATCTTTTGAACAGGGTAACAATATT 5631
Oy      283   ttggtgcttaatgycacataatgcataaacattcatcaatccccattatgccaagagc 342
Db      5630  ATTGGTGTATTTCCTCAAGATTTCCTCTCGTATTAATTGAAATCAAGCATGAAGTTCCTGGC 5571
Oy      343   aaattagcacttgcctaagactaaagtatgctt 375
Db      5570  AAAAAAGGAATTACTTTCTTCCAAAATATTATTTT 5538

```

Search completed: March 5, 2001, 06:22:52
Job time: 18038 sec

Qy 163 tatgtatlttatgcaggtataacttgcaccaaatttggcgtagaaaccgaatttgc 222

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 5, 2001, 01:54:04 ; Search time 79.21 Seconds
(without alignments)
2575.241 Million cell updates/sec

Title: US-09-164-714-6

Perfect score: 543
Sequence: 1 atgaacttaaaacact.....gcgcattgcttttaa 543

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N_Geneseq_36.*
1: /cgnl_8/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /cgnl_8/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /cgnl_8/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /cgnl_8/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /cgnl_8/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /cgnl_8/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /cgnl_8/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /cgnl_8/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /cgnl_8/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /cgnl_8/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /cgnl_8/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /cgnl_8/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /cgnl_8/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /cgnl_8/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /cgnl_8/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /cgnl_8/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /cgnl_8/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /cgnl_8/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /cgnl_8/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /cgnl_8/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /cgnl_8/gcgdata/geneseq/geneseqn/NA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	543	100.0	543	21	DNA encoding the O
2	71.2	13.1	80	21	Nucleotide sequence
3	41	7.6	1762	16	B31 Osp-A/antigen
4	40.4	7.4	18613	18	Staphylococcus aur
5	39.4	7.3	846	17	OspB soluble varia
6	39.4	7.3	891	16	OspB variant #1 co
7	39.4	7.3	891	17	B. burgdorferi str
8	39.4	7.3	1141	16	OspB variant #1 co
9	39.4	7.3	1180	16	B31 outer surface
10	39.4	7.3	1324	16	B31 outer surface
11	39.4	7.3	1363	16	B31 outer surface
12	39.4	7.3	1915	11	OspA and OspB-enco

13	39.4	7.3	1916	19	V00681	Borrelia burgdorfe
14	39.4	7.3	5898	19	X00677	DNA plasmid VR2211
15	39.4	7.3	53585	20	X20251	Borrelia burgdorfe
16	38.4	7.1	3323	18	T84223	DNA encoding a glu
17	38.4	7.1	3323	19	V53546	DNA encoding 2 Stra
18	36.4	6.7	400	18	V78384	Staphylococcus aur
19	35.6	6.6	3794	19	V00680	Plasmid pRR46 con
20	35.4	6.5	1214	18	V74424	Staphylococcus aur
21	35	6.4	19446	19	V52184	Streptococcus pneu
22	34.8	6.4	45	21	A12599	PCR primer used to
23	34.6	6.4	1349	19	X14172	H. pylori GHP0 639
24	34	6.3	4373	18	T72719	Pyruvate:Flavodoxi
25	33.4	6.2	4023	20	X13023	Enterococcus faeca
26	33.2	6.1	1015	14	O40695	B. burgdorferi str
27	32.4	6.0	2455	19	V26363	Moraxella catarrha
28	32.4	6.0	2896	13	O24135	Vector comprising
29	32	5.9	40	21	A12603	PCR primer used to
30	31.8	5.9	730	19	T98768	DNA encoding a S.
31	31.8	5.9	2531	19	V42988	Streptococcus jann
32	31.4	5.8	1664976	19	V21209	Methanococcus jann
33	31.2	5.7	584	19	V62140	HSV-2 strain SB5 C
34	31	5.7	2696	14	O41061	Ap Serotype 7 60KD
35	31	5.7	12438	20	X13110	Enterococcus faeca
36	31	5.7	235033	19	V57926	Hereditary haemoch
37	31	5.7	237326	19	V57903	Hereditary haemoch
38	30.8	5.7	2407	19	V26361	Moraxella catarrha
39	30.8	5.7	2520	21	Z38336	Moraxella catarrha
40	30.8	5.7	2520	21	Z35587	M. catarrhalis COP
41	30.8	5.7	2520	21	Z35588	M. catarrhalis COP
42	30.8	5.7	3762	18	T73217	AprXCA gene. ActI
43	30.8	5.7	3762	21	T86584	A. pleuropneumonia
44	30.8	5.7	4042	16	T04132	Helicobacter pylor
45	30.8	5.7	7528	20	X12992	Enterococcus faeca

ALIGNMENTS

RESULT 1
ID A12591
ID A12591 standard; DNA; 543 BP.
XX AC A12591;
XX DT 25-JUL-2000 (first entry)
DE DNA encoding the outer membrane protein 21 of strain ATCC49143.
XX XX
KW Outer membrane protein 21; OMP21; strain ATCC49143; bacterial infection;
KW ctitis media; respiratory infection; sinusitis; pneumonia; immunisation;
KW ds.
XX OS Moraxella catarrhalis.
XX FH Key
XX FT 1..543 location/Qualifiers
XX FT CDS
XX FT /*tag= a
XX FT /product= "outer membrane protein 21"
XX FT /transl_except= (pos: 211..213, aa: Ale)
XX PN WO200018910-A1.
XX PD 06-APR-2000.
XX PF 01-OCT-1999; 99WO-US22918.
XX PR 01-OCT-1998; 98US-0164714.
XX PA (ANTE-) ANTEX BIOLOGICS INC.
XX PI Tucker K, Tillmann UF;
XX DR WPI; 2000-293149/25.

DR P-PSDB: 184612.
 XR
 PR Isolated outer membrane protein from a *Moraxella catarrhalis* strain
 PT used for diagnosis treatment and prevention of disease caused by *M.*
 PT *catarrhalis* e.g. pneumonia, otitis media and respiratory infections -
 XX
 PS
 XX Claim 11; Fig 3; 108bp; English.
 CC The present sequence encodes an outer membrane protein 21 (OMP21) of
 CC *Moraxella catarrhalis* strain ATCC49143. The OMP21 protein has an
 CC apparent molecular weight of 16-20 kD as determined by sodium
 CC dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE). OMP21,
 CC its nucleic acids and antibodies can be used in prophylactic and
 CC therapeutic compositions for treating a *M. catarrhalis* bacterial
 CC infection, otitis media, respiratory infections, sinusitis and
 CC pneumonia. They are useful as reagents for the clinical or medical
 CC diagnosis of *M. catarrhalis* infections and for scientific research on
 CC the properties of pathogenicity, virulence and infectivity of
 CC *M. catarrhalis* and host defence mechanisms. The antibodies, particularly
 CC those that are cytotoxic may be used in passive immunisation to prevent
 CC or attenuate *M. catarrhalis* infections of animals e.g. humans.
 CC
 XQ Sequence 543 BP; 165 A; 108 C; 118 G; 152 T; 0 other;

Query Match	100.0%	Score 543;	DB 21;	Length 543;
Best Local Similarity	100.0%;	Pred. No. 4.8e-156;		
Matches 543; Conservative	0;	Mismatches	0;	Gaps 0

OY	1	atgaaacctttaaataaacacactatgagcgatctacagctcttcggtattatggcgatgagatgct	60
Db	1	atgaaacctttaaataaacacactatgagcgatcttcagctcttcggtattatggcgatgagatgct	60
OY	61	aagccttcacatcagctatctgacatcttcgcagatgacatgacacacatgcttgggtccaaatc	120
Db	61	aagccttcacatcagctatgacatcttcgcagatgacatgacacacatgcttgggtccaaatc	120
OY	121	ggctaaagtacagcgccaagcaaatcacgcgttaagaacaacgctatgattatctatgcagt	180
Db	121	ggctaaagtacagcgccaagcaaatcacgcgttaagaacaacgctatgattatctatgcagt	180
OY	181	tataactttgcacaaattttgtggtatgaaccccggaattttgtgtcttaagcgccaaagaa	240
Db	181	tataactttgcacaaattttgtggtatgaaccccggaattttgtgtcttaagcgccaaagaa	240
OY	241	tttaatgcagcgcgagatcctctgtataaagggtgagatgtagatcttcttggtgttatgtgaca	300
Db	241	tttaatgcagcgcgagatcctctgtataaagggtgagatgtagatcttcttggtgttatgtgaca	300
OY	301	tatcgcataacttcatacnaatlaccccatltaatgtgccaaaggccaatbaggcatgtgtaag	360
Db	301	tatcgcataacttcatacnaatlaccccatltaatgtgccaaaggccaatbaggcatgtgtaag	360
OY	361	actaaagtatagatgttaccacgcgttaatgtgcaactacatacctcaacaaagaagcacaacc	420
Db	361	actaaagtatagatgttaccacgcgttaatgtgcaactacatacctcaacaaagaagcacaacc	420
OY	421	agcctagcagcggtgtctgtgtgtttttgtgcgtltaaaccatltagcaaatgtggcgcttgaagca	480
Db	421	agcctagcagcggtgtctgtgtgtttttgtgcgtltaaaccatltagcaaatgtggcgcttgaagca	480
OY	481	agctacaacatcatcatcaagaagatgccaatgcaatlaagtttggcgctcatgtgacctt	540
Db	481	agctacaacatcatcatcaagaagatgccaatgcaatlaagtttggcgctcatgtgacctt	540
OY	541	taa 543	
Db	541	taa 543	

RESULT	2
A12605	
ID	A12605 standard; DNA; 80 BP.

XX A12605.
XX
XX 25-JUL-2000 (first entry)
XX
XX
XX Nucleotide sequence of a fragment of outer membrane protein 21 DNA.
XX
XX
XX Outer membrane protein 21; OMP21; strain ATCC49143; bacterial infection;
XX Otitis media; respiratory infection; sinusitis; pneumonia; Immunisation;
XX ss.
XX
XX Moraxella catarrhalis.
XX
XX WO200018910-A1.
XX
XX 06-APR-2000.
XX
XX 01-OCT-1999; 99WO-US22918.
XX
XX 01-OCT-1998; 98US-0164714.
XX
XX (ANTE-) ANTEX BIOLOGICS INC.
XX
XX Tucker K, Tillmann UF;
XX
XX WPI: 2000-293149/25.
XX
XX Isolated outer membrane protein from a Moraxella catarrhalis strain
XX used for diagnosis treatment and prevention of disease caused by M.
XX catarrhalis e.g. pneumonia, otitis media and respiratory infections -
XX
XX Example 8; Page 105; 108pp; English.
XX
XX The present sequence represents a PCR amplified fragment of the outer
XX membrane protein 21 (OMP21) gene from a Moraxella catarrhalis strain.
XX The OMP21 protein has an apparent molecular weight of 16-20 kD as
XX determined by sodium dodecylsulfate polyacrylamide gel electrophoresis
XX (SDS-PAGE). OMP21, its nucleic acids and antibodies can be used in
XX prophylactic and therapeutic compositions for treating a M. catarrhalis
XX bacterial infection, otitis media, respiratory infections, sinusitis
XX and pneumonia. They are useful as reagents for the clinical or medical
XX diagnosis of M. catarrhalis infections and for scientific research on
XX the properties of pathogenicity, virulence and infectivity of
XX M. catarrhalis and host defence mechanisms. The antibodies, particularly
XX those that are cytotoxic may be used in passive immunisation to prevent
XX or attenuate M. catarrhalis infections of animals e.g. humans.
XX
XX Sequence 80 BP; 26 A; 20 C; 18 G; 13 T; 3 other;

[illegible]

CC	and corresponding polynucleotides. Fusion proteins (i.e. the	
CC	above protein) comprising 2 or more antigenic Borrelia polypeptides,	
CC	that do not naturally occur in the same protein, can be used in the	
CC	treatment and diagnosis of Borrelia infections, i.e. as a vaccine	
CC	against Lyme borreliosis, in immunodiagnostic assays to detect	
CC	anti-Borrelia antibodies or to measure T-cell reactivity.	
XX		
S0	Sequence 1141 BP; 449 A; 206 C; 221 G; 265 T; 0 other;	
	Query Match	7.3%; Score 39.4; DB 16; Length 1141;
	Best Local Similarity	40.8%; Pred. No. 0.018;
	Matches 106; Conservative 0; Mismatches 111; Indels 0; Gaps 0;	
OY	33 agcttccttgatattggagatgagtgctcaagcgcgcacatcgcatggcaattctgcga	92
DB	114 agactcagtgctccttgatttaattgnaaataatttggtaagcaagaataaalaagctc	173
OY	93 tgcataaccctatgtgtgtgcacaaatltgtccaagtacgacgccaagcaaatcaacggtaa	152
DB	174 cggcaaatatgatattaagaagcaacacatgatcaggttgaaacttaaggaaacttcgataa	233
OY	153 gaacacgcgtatagtgatattatgcagggtatcaacttgcacaaaatttggcgtagaacc	212
DB	234 aaacatatgtcttggaacccttgaagcttgaagcctgcacaaagtaaaagtaaaattaac	293
OY	213 cgaatttgctgcagacgccaagaagaaatttaagca	249
DB	294 agttctgcgtatttaaacacagtaaccttagaagca	330
RESULT	9	
ID	Q90735	
XX	Q90735 standard; DNM; 1180 BP.	
XX	Q90735:	
AC		
XX	30-JUL-1996 (first entry)	
DE	B31 outer surface protein (Osp-A)/antigen P41 (122-234) fusion gene.	
XX		
KW	Strain: B31; antigen; antigenic domain; protein;	
KW	treatment; diagnosis; infection; vaccine; Lyme borreliosis;	
KW	immunodiagnostic assay; antibody; T-cell reactivity;	
KW	outer surface protein; Osp-A; antigen P41; fusion; ds.	
XX		
OS	Borrelia burgdorferi.	
XX		
XX	Key	Location/Qualifiers
FT	Mat_peptide	1..1180
FT		/*tag= a
XX		
XX	W09512676-A1.	
PD		
PD	11-MAY-1995.	
XX		
PE	27-OCT-1994;	94WO-US12352.
XX		
XX	29-APR-1994;	94US-0235836.
PR	01-NOV-1993;	93US-0148191.
XX		
PA	(ASU-) ASSOC UNIVERSITIES INC.	
XX		
PI	Dunn JI, Luft BJ;	
XX		
XX	WPI: 1995-215034/28.	
DR	P-PSDB: R75742.	
XX		
FT	Chimeric protein comprising 2 or more antigenic Borrelia	
FT	polypeptide(s) - useful in a vaccine against Lyme borreliosis and in	
XX	immunodiagnostic assays	
XX		
S5	Claim 43: F1q33: 200pp: Enq1sh.	

XX The present sequence encodes a fusion protein comprising outer
CC surface protein A (OspA) and antigen P41 (122-234) sequences, from
CC the B. burgdorferi strain B31, using chemical or enzymatic methods,
CC peptide fragments of the parent proteins were prepd., and analysed
CC by western blot to assess their ability to bind different
CC anti-OspA/P41 monoclonal antibodies. The information obtd. was used
CC to locate antigenic domains in the proteins, the epitopes of which
CC were mapped with the aid of site directed mutagenesis. Identical
CC analyses were performed on a selection of antigens purified from a
CC variety of B. burgdorferi strains, the results from which were
CC utilised in the prepn. of a pool of antigenic Borrelia polypeptides,
CC and corresponding polynucleotides. Fusion proteins (i.e. the
CC above protein) comprising 2 or more antigenic Borrelia polypeptides,
CC that do not naturally occur in the same protein, can be used in the
CC treatment and diagnosis of Borrelia infections, i.e. as a vaccine
CC against Lyme borreliosis, in immunodiagnostic assays to detect
CC anti-Borrelia antibodies, or to measure T-cell reactivity.
XX
XX Sequence 1180 bp; 466 A; 215 C; 224 G; 275 T; 0 other:

PA (ASU-) ASSOC UNIVERSITIES INC.
XX
PI Dunn JJ, Luft BJ:
XX
DR WPI; 1995-215034/28.
DR P-PSDB; R75745.
XX
XX Chimeric protein comprising 2 or more antigenic Borrelia
PT polypeptide(s) - useful in a vaccine against Lyme borreliosis and Ir
PT immuno:diagnostic assays
XX
XX
PS Claim 43; Fig 36; 200pp; English.
XX

Query Match	7.3%	Score 39.4;	DB 16;	Length 1180;
Best Local Similarity	48.8%;	Pred. No. 0.019;		
Matches 106;	Conservative 0;	Mismatches 111;	Indels 0;	Gaps 0;

[illegible]

```

OY      153 gaacacgcctatgttatattatgcaggttataacttcaaccaaaatttggcgtagaacc 212
          |||| | |||| | ||||| || |||| | | |
Db      234 aaacatgttcttgaacccttgaagttcaagcctgcacaagtgtaanaatlaaac 293

```

y 213 cgaattctgtgttcagacgcgaagaatttaatgca 249
 | | | | | | | | | | | | | | | | | |
 Db 294 agttctgtatttaacacagtaacctagaagca 330

RESULT	ID
Q90738	Q90738 standard; DNA; 1324 BP.

DT	30-JUL-1996 (first entry)
XX	
DE	B31 outer surface protein (Osp-A)/antigen P41 (140-235) fusion gene
XX	
KW	Strain; B31; antigen; antigenic domain; protein;
KW	treatment; diagnosis; infection; vaccine; Lyme borreliosis;
KW	immunodiagnostic assay; antibody; T-cell reactivity;
KW	outer surface protein; Osp-A; antigen P41; fusion; ds.
XX	
OS	<i>Borrelia burgdorferi</i> .

Protein	Key	Location/Qualifiers
FT	Mat-peptide	1.1324
FT		/rtag= a
XX		
PN	W09512676-A1.	
XX		
PD	11-MAV-1995.	
XX		
PF	27-OCN-1994;	94MO-US12352.
XX		
PR	29-APR-1994;	94US-0235836.
XX	01-NOV-1993;	93US-0148191.
XX		

Sequence 1324 BP; 521 A; 232 C; 249 G; 322 T; 0 other,

Query Match	7.33;	Score 39.4;	DB 16;	Length 1324;
Best Local Similarity	48.88;	Pred. No. 0.02;		
Matches 106;	Conservative	0;	Mismatches 111;	Indels 0;
				Gaps
QY 33	agctctctgcatatgvcgatagatgcttaacgcgcgcacatcagcatatgccaattctgctga	92		
Db 114	agactcagtgctcttgctgtttaatgtaataaaatttttgtaagaacagaaanaaatagctc	172		
QY 93	tgctcaaaccccatgcttggtgtgcacaaatttggttcaagttagaagccgaacaaatcaacgtaa	155		
Db 174	cggcaaaatagtatttaagaagaacaacatgcatcaggttgaacttaaaagaaacttcgataa	233		
QY 153	gaacacgccttagatgattatgacaggtatataactttaccacaaaatttggcgtagaac	212		
Db 234	aaacaaatgcttcgtgaaccccttgaaaggttcaaaagcgtgacaaaggttaaaataaattaac	293		
QY 213	cgaaattgcttgcttcagagcgcgaagaatttaattgca	249		
Db 294	agttctcgtcatattaaacacagataaccttgaagca	330		

RESULT	11
Q90736	
ID	Q90736 standard; DNA; 1363 BP

DT	30-JUL-1996	(first entry)
XX		
DE	B31	outer surface protein (Osp-A)/antigen P41 (122-295) fusion gene
XX		
KW	Strain:	B31; antigen; antigenic domain; protein;
KW	treatment;	diagnosis; infection; vaccine; Lyme borreliosis;
KW	immunodiagnostic	assay; antibody; T-cell reactivity;
KW	outer surface protein;	Osp-A; antigen P41; fusion; ds.

FH	Key	Location/Qualifiers
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100


```
Db 1143 agactcagtcgtcttggtaattgtaataaatttltgttaagcaagaagaaaaatagctc 1202
Oy 93 tgcacaacctatggttggcacaattgttgcagtagcagcaagcaaatcaacggttaa 152
Db 1203 cggcaaatatgtatttaaggaagcaaatatcatcaggttgaacttaaggaaactccgataa 1262
Oy 153 gaacaccgtttagtattatgcaagttataactttgacccaatttggcgtagaacc 212
Db 1263 aaacaatggtctcggaaccttggaaggttcaaaagcctgcaagaagtaaatataac 1322
Oy 213 cgaattgtgtgttcagcagcgcaagaatttaatgca 249
Db 1323 agttctgctgatttaaacacagtaaccttaagaagca 1359

RESULT 13
V00681
ID V00681 standard; DNA; 1916 BP.
XX
AC V00681;
XX
DE 08-JUN-1998 (first entry)
XX
DE Borrelia burgdorferi ospB gene region.
XX
KW Borrelia; antigen; outer surface protein; OspB; Lyme disease;
KW DNA vaccine; immunisation; VR2210; VR2211; vector; ss.
XX
OS Borrelia burgdorferi strain B31.
XX
XX Key Location/Qualifiers
FH RBS 139..144
FT CDS /*tag= a
FT /*tag= 152..973
FT /*tag= b 152..199
FT sig_peptide /*tag= c
FT /*tag= 200..970
FT /*tag= d 973..979
FT RBS /*tag= e
FT CDS 983..1873
FT /*tag= f
FT sig_peptide /*tag= g
FT /*tag= 1028..1870
FT /*tag= h
FT mat_peptide
XX
XX WO9747197-A1.
XX
XX 18-DEC-1997.
XX
XX 03-JUN-1997; 97WO-US09439.
XX
XX 14-JUN-1996; 96US-0663998.
XX
XX (INMR ) PASTEUR MERIEUX SERUMS & VACCINS.
XX (UYTE-) UNIV TEXAS HEALTH SCI CENT SAN ANTONIO.
XX (VICA-) VICAL INC.
XX
XX Barbour AG, Garner KR, Huebner RC, Liang X, Luke CJ,
XX PI Norman JA;
XX
XX WPI: 1998-051938/05.
XX DR P-PSDB; W26771 AND W37131.
XX
XX Plasmid for expressing Borrelia antigen in eukaryotic cells - used
XX PT as vaccines to protect against Lyme disease and for production of
XX PT antigens, themselves used in vaccines
XX
XX PS Example 1; Fig 7A-H; 80pp; English.
XX
XX This nucleotide sequence comprises the ospB and ospB gene region
```

```
CC of Borrelia burgdorferi. The ospA and ospB genes respectively
CC code for outer surface proteins (Osp) (see W37131) and OspB
CC (see W26771) of B. burgdorferi. The ospA and ospB genes have
CC been used in the construction of VR2210 (see V00676) and VR2211
CC (see V00677), DNA plasmids suitable for transfection of e.g. baby
CC hamster kidney and human melanoma cells. Plasmid DNA encoding at
CC least one Borrelia genospecies antigen is disclosed and claimed.
CC The genospecies may be B. burgdorferi, Borrelia garinii and/or
CC Borrelia afzelii. The antigen can be OspA and/or OspB and/or OspC.
CC The plasmid is used in vaccines to elicit a protective immune
CC response (both humoral and cellular) in hosts, human or animal,
CC susceptible to Lyme disease (claimed). It can also be used for the
CC production of a Borrelia antigen in vitro in eukaryotic cells
CC (claimed), and these antigens used as vaccines or immunogenic
CC compositions, or to produce monoclonal antibodies.
XX
SQ Sequence 1916 BP; 810 A; 280 C; 337 G; 489 T; 0 other;

Query Match 7.3%; Score 39.4; DB 19; Length 1916;
Best Local Similarity 48.8%; Pred. No. 0.023;
Matches 106; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

Oy 33 agcttcttgatttggcgaatgagtcacgcctgcacatcagctatgcaattctgctga 92
Db 1144 agactcagtcgtcttggtaattgtaataaatttltgttaagcaagaagaaaaatagctc 1203
Oy 93 tgcacaacctatggttggcacaattgttgcagtagcagcaagcaaatcaacggttaa 152
Db 1204 cggcaaatatgtatttaaggaagcaaatatcatcaggttgaacttaaggaaactccgataa 1263
Oy 153 gaacaccgtttagtattatgcaagttataactttgacccaatttggcgtagaacc 212
Db 1264 aaacaatggtctcggaaccttggaaggttcaaaagcctgcaagaagtaaatataac 1323
Oy 213 cgaattgtgtgttcagcagcgcaagaatttaatgca 249
Db 1324 agttctgctgatttaaacacagtaaccttaagaagca 1360

RESULT 14
V00677
ID V00677 standard; DNA; 5898 BP.
XX
AC V00677;
XX
DE 08-JUN-1998 (first entry)
XX
XX DNA plasmid VR2211 containing Borrelia burgdorferi OspB gene.
XX
XX Borrelia; antigen; outer surface protein A; OspB; Lyme disease;
XX KW DNA vaccine; immunisation; VR2211; vector; ss.
XX
XX OS Chimeric - Homo sapiens.
XX
XX FH Key Location/Qualifiers
FT primer_bind /*tag= a
FT /*tag= 2..19
FT /*note= "binding site of forward primer used to
FT amplify and insert OspB coding sequence"
FT primer_bind 835..852
FT /*tag= b
FT /*note= "binding site of reverse primer used to
FT amplify and insert OspB coding sequence"
XX
XX WO9747197-A1.
XX
XX 18-DEC-1997.
XX
XX 03-JUN-1997; 97WO-US09439.
XX
XX 14-JUN-1996; 96US-0663998.
XX
```

```

XX (INMR) PASTEUR MERIEUX SERUMS & VACCINS.
PA (UTRE-) UNIV TEXAS HEALTH SCI CENT SAN ANTONIO.
PA (VICA-) VICAL INC.
PI Barbour AG, Garner KR, Huebner RC, Liang X, Luke CJ,
PI Norman JA;
XX WPI: 1998-051938/05.
XX
XX Plasmid for expressing Borrelia antigen in eukaryotic cells - used
PT as vaccines to protect against Lyme disease and for production of
PT antigens, themselves used in vaccines
XX
XX Example 1: Fig 4A-D; 80pp; English.
XX
XX This is the nucleotide sequence of VR2211, a DNA plasmid suitable
CC for transfection of e.g. baby hamster kidney and human melanoma
CC cells. VR2211 contains DNA encoding the Borrelia burgdorferi
CC outer surface protein B (OspB). It was made by ligating: an OspB
CC fragment (see V00680) amplified from pTRH46 (see V00685-86); a
CC tissue plasminogen activator 5'-untranslated region/leader peptide
CC sequence amplified from nKCMintBL (see V00678); and pSt1- and
CC xbaI-digested plasmid VR1012. Plasmid DNA encoding at least one
CC Borrelia genespecies antigen is disclosed and claimed. The
CC genespecies may be B. burgdorferi, Borrelia garinii and/or
CC Borrelia afzelii. The antigen can be OspA and/or OspB and/or OSpC.
CC The plasmid is used in vaccines to elicit a protective immune
CC response (both humoral and cellular) in hosts, human or animal,
CC susceptible to Lyme disease (claimed). It can also be used for the
CC production of a Borrelia antigen in vitro in eukaryotic cells
CC (claimed), and these antigens used as vaccines or immunogenic
CC compositions, or to produce monoclonal antibodies.
XX
XX Sequence 5898 BP, 1643 A; 1402 C; 1380 G; 1473 T; 0 other;
SQ

```

Query Match 7.3%; Score 39.4; DB 19; Length 5898;
 Best Local Similarity 48.8%; Pred. No. 0.037; Indels 0; Gaps 0;
 Matches 106; Conservative 0; Mismatches 111;

```

OY 33 agctcttcgtatcgtgcgatgctgaacgtccatcagcatatgcaattcgtcga 92
DB 118 agaccacagtgctcgttcaatgtaataaatttttgaagcaaaataaataagc 177
OY 93 tgcataacctatgttgcgaataatggtcgaagtagcgcgaagcaaatcaacggtaa 152
DB 178 cggcaaatatgatttaagaacacaaatgataagttgaactlaaagaaacttcgataa 237
OY 153 gaacacgcgtatggtattatcaggtatcacttgcacaaattttggtcgtagaacc 212
DB 238 aaacaaatggtctcgaaaccttgaaagttcgaagcctgcgaagaaagtaaaataaac 297
OY 213 cgaattgtgtcagacgcgcaagaatttaatgca 249
DB 298 agttctgtgatttaaacacagtaaccttagaagca 334

```

RESULT 15
 X20251/c
 ID X20251 standard; DNA: 53585 BP.
 AC X20251;
 XX
 XX 04-MAY-1999 (first entry)
 XX
 XX Borrelia burgdorferi polynucleotide sequence #4.
 DE
 XX
 XX Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
 KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
 KW infection; diagnosis; characterisation; detection; ds.
 XX
 OS Borrelia burgdorferi.

```

XX WO9858943-A1.
PN
XX
XX 30-DEC-1998.
PD
XX
XX 18-JUN-1998; 98WO-US12764.
PF
XX
XX 03-SEP-1997; 97US-0057483.
PR 20-JUN-1997; 97US-0050359.
PR 22-JUL-1997; 97US-0053344.
PR 22-JUL-1997; 97US-0053377.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA (MEDI-) MEDIMUNE INC.
XX
XX Clayton R, Dougherty BA, Fraser C, Lathigra K, Smith HO;
PI White OR;
XX
XX WPI: 1999-081217/07.
XX
XX New isolated Borrelia burgdorferi nucleic acids - used to develop
PT products for the detection, diagnosis, characterisation, prevention
PT and therapy of infections, particularly Lyme disease
XX
XX Claim 1: Page 801-831; 1128pp; English.
XX
XX X20248 to X20402 represent polynucleotide sequences isolated from
CC Borrelia burgdorferi (Bb). Products derived from Bb can be used for
CC the detection, diagnosis, characterisation, prevention and therapy of
CC Bb infections, e.g. Lyme disease. They can also be used for the
CC production of biosynthetic products, e.g. enzymes. Borrelia belongs
CC to a family of mollic, spiral-shaped bacteria called Spirochetes.
CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and
CC endemic relapsing fever, and Lyme borreliosis, more commonly known as
CC Lyme disease.
XX
XX Sequence 53585 BP; 19235 A; 8081 C; 7001 G; 19266 T; 2 other;
SQ

```

Query Match 7.3%; Score 39.4; DB 20; Length 53585;
 Best Local Similarity 48.8%; Pred. No. 0.096; Indels 0; Gaps 0;
 Matches 106; Conservative 0; Mismatches 111;

```

OY 33 agctcttcgtatcgtgcgatgctgaacgtccatcagcatatgcaattcgtcga 92
DB 43176 AGACTCAGTGTCTTTGTTAATGTAATTAATTTTGTAAAGCAAAATAATAGCTC 43117
OY 93 tgcataacctatgttgcgaataatggtcgaagtagcgcgaagcaaatcaacggtaa 152
DB 43116 CCGCAAAATATGATTTTAAGCAACAATTAATGATCAAGTTGAACTTAAAGAACTTCGATAC 43057
OY 153 gaacacgcgtatggtattatcaggtatcacttgcacaaattttggtcgtagaacc 212
DB 43056 AAACAAATGTTCTCGAACCCTTGAAAGTTCAAGCTGACAGAGTAAGTAATAATTAAC 42997
OY 213 cgaattgtgtcagacgcgcaagaatttaatgca 249
DB 42996 AGTTCTGCTGATTTAAACACAGTAACCTTAAGAGCA 42960

```

Search completed: March 5, 2001, 06:20:48
 Job time: 16004 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 5, 2001, 01:28:06 ; Search time 62.98 Seconds
(without alignments)
1389.489 Million cell updates/sec

Title: US-09-164-714-6
Perfect score: 543
Sequence: 1 atgaatacttaaacact.....gcgtcattggttttaa 543

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 280836 seqs, 80580151 residues

Total number of hits satisfying chosen parameters: 561672

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgnl_7/ptodata/1/lna/5A.COMB.seq:*
2: /cgnl_7/ptodata/1/lna/5B.COMB.seq:*
3: /cgnl_7/ptodata/1/lna/6.COMB.seq:*
4: /cgnl_7/ptodata/1/lna/6CTUS.COMB.seq:*
5: /cgnl_7/ptodata/1/lna/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39.4	7.3	846	1	US-07-941-523-21
2	39.4	7.3	891	1	US-07-941-523-19
3	39.4	7.3	1959	1	US-08-137-175A-1
4	39.4	7.3	1959	3	US-08-479-017-1
5	39.4	7.3	5952	2	US-08-663-998-2
6	31	5.7	2696	1	US-07-961-522-1
7	31	5.7	2696	1	US-08-321-438-1
8	31	5.7	2696	1	US-08-321-978-1
9	31	5.7	2696	2	US-08-710-584-1
10	31	5.7	246240	2	US-08-724-394A-20
11	31	5.7	246240	2	US-08-724-394A-21
12	31	5.7	246240	2	US-08-724-394A-22
13	30.8	5.7	2520	2	US-08-450-351-1
14	30.8	5.7	2520	2	US-08-450-351-3
15	30.8	5.7	3762	3	US-08-772-270A-3
16	30.8	5.7	4042	3	US-08-200-232-1
17	30.8	5.7	4042	4	PCT-US95-02219-1
18	30.8	5.7	4042	4	PCT-US95-02219A-1
19	30.8	5.7	8370	2	US-08-488-706-1
20	30.6	5.6	2278	1	US-08-258-188-1
21	30.6	5.6	2278	1	US-08-526-813-1
22	30.6	5.6	2278	4	PCT-US95-08554-1
23	30.2	5.6	994	3	US-08-961-083-211
24	30.2	5.6	1033	3	US-08-961-083-191
25	30.2	5.6	152331	3	US-09-128-155-17
26	30.2	5.6	176373	3	US-09-128-155-17
27	29.8	5.5	1903	1	US-07-961-522-3
28	29.8	5.5	1903	1	US-08-217-438-3

29	29.8	5.5	1903	1	US-08-321-978-3	Sequence 3, Appl1
30	29.8	5.5	1903	2	US-08-710-584-3	Sequence 3, Appl1
31	29.4	5.4	660	2	US-08-625-377-1	Sequence 1, Appl1
32	29.4	5.4	3960	3	US-08-470-260-2	Sequence 2, Appl1
33	29.4	5.4	3960	3	US-08-471-491-2	Sequence 2, Appl1
34	29.4	5.4	3960	3	US-08-466-662-2	Sequence 2, Appl1
35	29.2	5.4	839	3	US-08-642-807A-30	Sequence 30, Appl1
36	29.2	5.4	1529	2	US-08-477-451-17	Sequence 17, Appl1
37	29.2	5.4	1529	2	US-08-477-451-21	Sequence 21, Appl1
38	29.2	5.4	19932	2	US-08-477-451-25	Sequence 25, Appl1
39	29	5.3	4140	3	US-08-894-731-2	Sequence 2, Appl1
40	28.8	5.3	2949	3	US-08-433-522A-1	Sequence 1, Appl1
41	28.8	5.3	2949	3	US-09-135-166-1	Sequence 1, Appl1
42	28.8	5.3	3294	1	US-08-409-995-1	Sequence 1, Appl1
43	28.6	5.3	3294	3	US-08-685-467-1	Sequence 1, Appl1
44	28.6	5.3	5319	1	US-08-159-927-1	Sequence 1, Appl1
45	28.4	5.2	1716	3	US-09-029-267-1	Sequence 1, Appl1

ALIGNMENTS

```
RESULT 1
US-07-941-523-21
; Sequence 21, Application US/07941523
; Patent No. 5571718
; GENERAL INFORMATION:
; APPLICANT: Dunn, John J
; APPLICANT: Barbour, Alan G
; TITLE OF INVENTION: Cloning and Expression of Borrelia
; NUMBER OF INVENTION: Lipoproteins
; CORRESPONDENCE ADDRESS: 24
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 01730
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/941,523
; FILING DATE: 19920908
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; TELEPHONE: (617) 861-6240
; TELECOMMUNICATION INFORMATION:
; REFERENCE/DOCKET NUMBER: BNL90-01A
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 846 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-07-941-523-21
```

Query Match 7.3% Score 39.4; DB 1; Length 846;
Best Local Similarity 48.8%; Pred. No. 0.0037;
Matches 106; Conservative 0; Mismatches 111; Indels 0; Caps 0;

33 agctcttcgtatctgagatgagtgcaacgcgcacatcgcatatgcaatctgcgga 92
DB 117 AGACTCAGCTCTTGTGTAATGTAATAATTTTGTATGACAAAGAAAAAATACCTC 176

NAME/KEY: CDS
LOCATION: 950..1840
OTHER INFORMATION: /product= "ospb"
US-08-137-175A-1

Query Match 7.3%; Score 39.4; DB 1; Length 1959;
Best Local Similarity 48.8%; Pred. No. 0.0053;
Matches 106; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

QY 33 agctcttcgttatgtgcatgagtgctaacgctgcacatgaagctatggaattctgtga 92
DB 1111 AGACTCAGTGTCTGTTTAAATGATTAATTTTGTAAAGCAAGAAAAAATAGCTC 1170
QY 93 tgcacaaccctatgtgtgbcacaaatttgctcaagtagacgcacaaatcaacggttaa 152
DB 1171 CGGCAATATGATTTAAGACACAAATTTGATCAGTTGAACCTTAAGAACTTCCGATTA 1230
QY 153 gaacaccctatgtatattatgaaggtataaacttgacaaaatttgcgtagaacc 212
DB 1231 AAACAATGTTCTGGAACCTTGAAAGGTTCAGACCTGACAGAGTAAGTAATAATTAAC 1290
QY 213 cgaatttgtgttcagagcccaagaatttaagca 249
DB 1291 AGTTTCGTGATTTAAACACAGTAACCTTAGAAGCA 1327

RESULT 4
US-08-479-017-1
Sequence 1, Application US/08479017
Patent No. 6143872

GENERAL INFORMATION:
APPLICANT: BARBOUR, Alan G.
APPLICANT: BERGSTROM, Sven
APPLICANT: HANSSON, Lennart
TITLE OF INVENTION: IMPROVEMENT IN BORRELIA BURGDORFERI AND
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROMBY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/479,017
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/137,175
FILING DATE: 26-OCT-1993
APPLICATION NUMBER: PCT/US92/08972
FILING DATE: 22-OCT-1992
ATTORNEY/AGENT INFORMATION:
NAME: COOPER, Iver P.
REGISTRATION NUMBER: 28,005
REFERENCE/DOCKET NUMBER: BARBOUR-1B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1959 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Borrelia burgdorferi
STRAIN: B31 (ATCC 35210)

FEATURE:
NAME/KEY: misc_feature
LOCATION: 123..142
OTHER INFORMATION: /function= "primer"

FEATURE:
NAME/KEY: misc_feature
LOCATION: 584..607
OTHER INFORMATION: /function= "primer"

FEATURE:
NAME/KEY: misc_feature
LOCATION: 776..794
OTHER INFORMATION: /function= "primer"

FEATURE:
NAME/KEY: misc_feature
LOCATION: 806..817
OTHER INFORMATION: /function= "primer"

NAME/KEY: CDS
LOCATION: 119..940
OTHER INFORMATION: /product= "ospa"

FEATURE:
NAME/KEY: CDS
LOCATION: 950..1840
OTHER INFORMATION: /product= "ospb"

Query Match 7.3%; Score 39.4; DB 3; Length 1959;
Best Local Similarity 48.8%; Pred. No. 0.0053;
Matches 106; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

QY 33 agctcttcgttatgtgcatgagtgctaacgctgcacatgaagctatggaattctgtga 92
DB 1111 AGACTCAGTGTCTGTTTAAATGATTAATTTTGTAAAGCAAGAAAAAATAGCTC 1170
QY 93 tgcacaaccctatgtgtgbcacaaatttgctcaagtagacgcacaaatcaacggttaa 152
DB 1171 CGGCAATATGATTTAAGACACAAATTTGATCAGTTGAACCTTAAGAACTTCCGATTA 1230
QY 153 gaacaccctatgtatattatgaaggtataaacttgacaaaatttgcgtagaacc 212
DB 1231 AAACAATGTTCTGGAACCTTGAAAGGTTCAGACCTGACAGAGTAAGTAATAATTAAC 1290
QY 213 cgaatttgtgttcagagcccaagaatttaagca 249
DB 1291 AGTTTCGTGATTTAAACACAGTAACCTTAGAAGCA 1327

RESULT 5
US-08-663-998-2
Sequence 2, Application US/08663998
Patent No. 5845946

GENERAL INFORMATION:
APPLICANT: HUEBNER, Robert C.
APPLICANT: NORMAN, Jon A.
APPLICANT: LIANG, Xiaowu
APPLICANT: CARNER, Kristin R.
APPLICANT: BARBOUR, Alan G.
APPLICANT: LUKE, Catherine J.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Curtis, Morris & Safford, P.C.
STREET: 530 Fifth Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/663,998
FILING DATE: 06-JUN-1996
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Kowalski, Thomas J.
REGISTRATION NUMBER: 32,147
REFERENCE/DOCKET NUMBER: 454,312-2440
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-840-3333
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 5952 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-663-998-2
```

```
Query Match 7.3%; Score 39.4; DB 2; Length 5952;
Best Local Similarity 48.8%; Pred. No. 0.0087;
Matches 106; Conservative 0; Mismatches 111; Indels 0; Gaps 0;
```

```
QY 33 agcttctcgtatagcgatgagtgtaacgtcgcacatcagctatggaatctgctga 92
DB 1980 AGAGCTAGAGTCTCTTTTAAGCTAATATAATTTTGTAGAGCAAGAAAATAATGCTC 2039
QY 93 tgcacaacctatgttgtgtgccaanaattgttcaagtagaagcccaagcaaatcaacgttaa 152
DB 2040 CGGCAATATGATTTAAGCAACCAATGATGATCGTTGAACCTTAAGGAACCTCGATTA 2099
QY 153 gaacaccgttatagatattatgcaagttataaacttgcacaaatcttggcgtaaac 212
DB 2100 AACAAGATGTTCTGGAACCTTGGAAGCTTCAAGCTGACAGACGTAAGTAATAATTAC 2159
QY 213 cgaatttgtgttcagacgcgaagcaaatltaatga 249
DB 2160 AGTTTCTGCTGATTTAACACAGTACCTTAGAGCA 2196
```

```
RESULT 6
US-07-961-522-1
Sequence 1, Application US/07961522
Patent No. 5417971
GENERAL INFORMATION:
APPLICANT: POTTER, ANDREW A.
APPLICANT: GERLACH, GERALD F.
APPLICANT: WILLSON, PHILIP J.
APPLICANT: ROSSI-CAMPOS, AMALIA
TITLE OF INVENTION: VACCINES FOR ACTINOBACILLUS
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: ROBERTA L. ROBINS
STREET: 635 BRYANT STREET
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94301
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/961,522
FILING DATE: 19921015
```

```
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: ROBINS, ROBERTA L.
REGISTRATION NUMBER: 33,208
REFERENCE/DOCKET NUMBER: 9000-0015.20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 617-8999
TELEFAX: (415) 327-3231
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2696 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 333..1973
US-07-961-522-1
```

```
Query Match 5.7%; Score 31; DB 1; Length 2696;
Best Local Similarity 47.0%; Pred. No. 2.6;
Matches 131; Conservative 0; Mismatches 145; Indels 3; Gaps 1;
```

```
QY 235 aaagaatttaatgcaaggcgtgaagtcclglaaaagtgatgaagtccttggcttat 294
DB 1662 ACAGAAATTGATGCTGATTTGGCAACAAAATTTAAGTGTAAAGTTATTTGCA1AAAAAC 1721
QY 295 ggcacatacgcataactatcacaacacccatllatgccaaggcaaatgaagcatt 354
DB 1722 GGGTGAATCTCTGTGTTTACCGTAGATG:AAAATTGATGTAATGTTTACTGGCAAA 1781
QY 355 gctaaagctaaagtagatgttaccagcgttaatgcaactacatcacaacaaagcgac 414
DB 1782 GCTAANAACCTCAGATGAAGGCTCGCTAGATTCAAGTACAGTTACAGTTATGAGATG 1841
QY 415 aaacagccttagcaggcggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 474
DB 1842 AATTTAAGAGTAGTACAGTACGTAGTGTGCTGTATGTGTCAACGCGCA--CGAGAGCTT 1898
QY 475 gaagcagctaacatcatcatcagaagatgcgaatga 513
DB 1899 GCGGCAATTCACCATTAATCAGAAATGGCAGTGA 1937
```

```
RESULT 7
US-08-217-438-1
Sequence 1, Application US/08217438
Patent No. 5521072
GENERAL INFORMATION:
APPLICANT: Potter, Andrew A.
APPLICANT: Gerlach, Gerald F.
APPLICANT: Willson, Philip J.
APPLICANT: Rossi-Campos, Amalia
TITLE OF INVENTION: ACTINOBACILLUS PLEUROPEINOMIAE
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Reed & Robins
STREET: 635 Bryant Street
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94301
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/217,438
FILING DATE: 22-MAR-1994
```

```
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: ROBINS, ROBERTA L.
REGISTRATION NUMBER: 33,208
REFERENCE/DOCKET NUMBER: 9001-0015.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-617-8999
TELEFAX: 415-327-3231
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2696 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 333..1973
US-08-217-438-1
```

```
Query Match          5.7% Score 31; DB 1; Length 2696;
Best Local Similarity 47.0% Pred. No. 2.6;
Matches 131; Conservative 0; Mismatches 145; Indels 3; Gaps 1;
```

```
QY 235 aaagaatttaatgcaggcgtagctcctgtaaaagtgatggaagctcttggcgcttat 294
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1662 ACAGAAATTTGATGTTGATTTTGGCAACAAAATTAACTGTAAGTTATTTGATAAANAAC 1721
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 295 ggcacatatgcctataaactcaatcaataaccatttattgccaaggccaattagcatt 354
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1722 GGGTAAATCTCTGTTTACCGTAGATGCAAAAATTGATGTAATGTTTACTGCGCAA 1781
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 355 gctaaagcctaagtagatgtttaccagcgtaagcctaactacatactcaaaagaagcgc 414
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1782 GCTAAACCTCAGATGAAGGCTTGCTCTAGATTCAAGTTACGTTATGAGAAATGTG 1841
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 415 aaaaccagcctagcaggcggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 474
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1842 AAATTTAACGATGTAGACAGTTAGTGTGCGTCTATGTGTCCAAACGGCA---CCAGACCTT 1898
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 475 gaagcaagcctaacatctatctacagaagatgcacatgca 513
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1899 GCGGACAAATTCACCATTAATCAGAAAATGCGAGTGTA 1937
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 8
US-08-321-978-1
Sequence 1, Application US/08321978
Patent No. 5801018
```

```
GENERAL INFORMATION:
APPLICANT: POTTER, ANDREW A.
APPLICANT: GERLACH, GERALD F.
APPLICANT: WILSON, PHILIP J.
APPLICANT: ROSSI-CAMPOS, AMALIA
TITLE OF INVENTION: VACCINES FOR ACTINOBACILLUS
TITLE OF INVENTION: PLEUROPNEUMONIAE
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESS: ROBERTA L. ROBINS
STREET: 285 HAMILTON AVE, SUITE 200
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94301
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/321,978
FILING DATE: 12-OCT-1994
```

```
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: ROBINS, ROBERTA L.
REGISTRATION NUMBER: 33,208
REFERENCE/DOCKET NUMBER: 9001-0015.02
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 327-3400
TELEFAX: (415) 327-3231
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2696 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 333..1973
US-08-321-978-1
```

```
Query Match          5.7% Score 31; DB 1; Length 2696;
Best Local Similarity 47.0% Pred. No. 2.6;
Matches 131; Conservative 0; Mismatches 145; Indels 3; Gaps 1;
```

```
QY 235 aaagaatttaatgcaggcgtagctcctgtaaaagtgatggaagctcttggcgcttat 294
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1662 ACAGAAATTTGATGTTGATTTTGGCAACAAAATTAACTGTAAGTTATTTGATAAANAAC 1721
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 295 ggcacatatgcctataaactcaatcaataaccatttattgccaaggccaattagcatt 354
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1722 GGGTAAATCTCTGTTTACCGTAGATGCAAAAATTGATGTAATGTTTACTGCGCAA 1781
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 355 gctaaagcctaagtagatgtttaccagcgtaagcctaactacatactcaaaagaagcgc 414
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1782 GCTAAACCTCAGATGAAGGCTTGCTCTAGATTCAAGTTACGTTATGAGAAATGTG 1841
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 415 aaaaccagcctagcaggcggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 474
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1842 AAATTTAACGATGTAGACAGTTAGTGTGCGTCTATGTGTCCAAACGGCA---CCAGACCTT 1898
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 475 gaagcaagcctaacatctatctacagaagatgcacatgca 513
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1899 GCGGACAAATTCACCATTAATCAGAAAATGCGAGTGTA 1937
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 9
US-08-710-584-1
Sequence 1, Application US/08710584
Patent No. 5876725
```

```
GENERAL INFORMATION:
APPLICANT: POTTER, ANDREW A.
APPLICANT: GERLACH, GERALD F.
APPLICANT: WILSON, PHILIP J.
APPLICANT: ROSSI-CAMPOS, AMALIA
TITLE OF INVENTION: VACCINES FOR ACTINOBACILLUS
TITLE OF INVENTION: PLEUROPNEUMONIAE
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESS: ROBERTA L. ROBINS
STREET: 285 HAMILTON AVE, SUITE 200
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94301
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/710,584
FILING DATE:
```

```

? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/321,978
? FILING DATE: 12-OCT-1994
? ATTORNEY/AGENT INFORMATION:
? NAME: ROBINS, ROBERTA L.
? REGISTRATION NUMBER: 33,208
? REFERENCE/DOCKET NUMBER: 9001-0015.02
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (415) 327-3400
? TELEFAX: (415) 327-3231
? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 2696 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 333..1973
? US-08-710-584-1

```

```

1  COMPUTER: IBM PC compatible
2  OPERATING SYSTEM: PC-DOS/MS-DOS
3  SOFTWARE: Patent in Release #1.0, Version #1.30
4
5  CURRENT APPLICATION DATA:
6  APPLICATION NUMBER: US/08/724,394A
7  FILING DATE: 01-OCT-1996
8  CLASSIFICATION: 536
9
10 ATTORNEY/AGENT INFORMATION:
11 NAME: Filts, Renee A.
12
13 REGISTRATION NUMBER: 35,136
14 REFERENCE/DOCKET NUMBER: 017957-000100
15
16 TELECOMMUNICATION INFORMATION:
17 TELEPHONE: 415-576-0200
18 TELEFAX: 415-576-0300
19
20 INFORMATION FOR SEQ ID NO: 20:
21 SEQUENCE CHARACTERISTICS:
22 LENGTH: 246240 base pairs
23 TYPE: nucleic acid
24 STRANDEDNESS: not relevant
25 TOPOLOGY: not relevant
26
27 MOLECULE TYPE: cdna
28
29 FEATURE:
30

```

Query Match	5.7%	Score	31	DB	2	Length	2696
Best Local Similarity	47.0%	Pred. No.	2.6				
Matches	131	Conservative	0	Mismatches	145	Indels	3
						Gaps	1

OTHER INFORMATION: /note- "HLA-H.CONFIG"
US-08-724-394A-20

[illegible]

Query Match	5.7%	Score 31	DB 2	Length 246240
Best Local Similarity	51.0%	Pred. NO. 19		
Matches 73	Conservative	0	Mismatches	70
			Indels	0
			Gaps	0
QY	16	aaacattatgagatgaatcaagctctcttcglatatgtgcatgtagtgaatgaatgcacatcaagc	75	
Db	75473	ACACACTTGCAACATCAACACTTTCAGATTATTTCCGTGGCTGACATACCTCCACAI	IATC	75532
QY	76	tatggcaattctctgcgatgcgtcaaccctatglttgatgcgaataatcggcctaagaagaagcc	135	
Db	75533	AGAGCCACATTTCTTTTATGGAATATTTTAGTGTTGTGCAAAAGTAAATGGCGGCTTTGGC		75592
QY	136	aagcaaatcaacaggtgaagaacac	158	
Db	75593	ATTAAAGTAAAGGCAAAAACACAC		75615
RESULT	11			
OS-08-724-394A-21				

```

10 RESULT 10
11 US-08-724-394A-20
12 : Sequence 20 Application US/08724394A
13 : Patent No. 5872237
14 :
15 : GENERAL INFORMATION:
16 : APPLICANT: Feder, John N.
17 : APPLICANT: Krommal, Gregory S.
18 : APPLICANT: Lauert, Peter M.
19 : APPLICANT: Ruddy, David A.
20 : APPLICANT: Thomas, Winston
21 : APPLICANT: Tsuchinashi, Zenta
22 : APPLICANT: Wolff, Roger K.
23 : TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
24 : TITLE OF INVENTION: Sequences and Antibodies Thereeto
25 : NUMBER OF SEQUENCES: 31
26 :
27 : CORRESPONDENCE ADDRESS:
28 : ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
29 : STREET: Two Embarcadero Center, 8th Floor
30 : CITY: San Francisco
31 : STATE: CA
32 : COUNTRY: USA
33 : ZIP: 94111-3834
34 :
35 : COMPUTER READABLE FORM:
36 : MEDIUM TYPE: Floppy disk

```

GENERAL INFORMATION:
APPLICANT: Feder, John N.
APPLICANT: Krommal, Gregory S.
APPLICANT: Iauer, Peter M.
APPLICANT: Ruddy, David A.
APPLICANT: Thomas, Winston
APPLICANT: Tsuchinashi, Zenta
APPLICANT: Wolff, Roger K.
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237c1
TITLE OF INVENTION: Sequences and Antibodies Thereo
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

```
APPLICATION NUMBER: US/08/724,394A
FILING DATE: 01-OCT-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Fitts, Renee A.
REGISTRATION NUMBER: 35,136
REFERENCE/DOCKET NUMBER: 017957-000100
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-576-0200
TELEFAX: 415-576-0300
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 246240 base pairs
TYPE: nucleic acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..246240
OTHER INFORMATION: /note= "HLA-H.CONFIG"
US-08-724-394A-21
```

```
Query Match          5.7%  Score 31:  DB 2:  Length 246240:
Best Local Similarity 51.0%  Pred. No. 19:
Matches 73:  Conservative 0;  Mismatches 70;  Indels 0;  Gaps 0;
```

```
OY 16 acacattgacgagtcacgcttcctcgtatgtgagtgagtgacgctgcacacg 75
    ||||| ||| ||| ||| ||||| ||| ||||| ||| ||| ||| ||| |||
DB 75473 ACACACTTGCACTCAACAACTTCAGATTATTTCCGCTGCTGACTCACTCCACATTATTC 75532
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 76 tatggcaattctgctgatactcaaccctatgtgtgtgcacaaatgtgcaagtaagcgc 135
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 75533 AGAGCCACTTCCTTTATGGAATATTATGCTTGTCGCAAAAGTATTGCGGCTTTGCC 75592
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 136 aagcaaatcaacgyltaagaacac 158
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 75593 ATTAAAGTAAAGCAAAAACACAC 75615
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 12
US-08-724-394A-22
Sequence 22, Application US/08724394A
Patent No. 5872237
GENERAL INFORMATION:
APPLICANT: Feder, John N.
APPLICANT: Krommal, Gregory S.
APPLICANT: Lauer, Peter M.
APPLICANT: Ruddy, David A.
APPLICANT: Thomas, Winston
APPLICANT: Tsuchihashi, Zenta
APPLICANT: Wolff, Roger K.
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
TITLE OF INVENTION: Sequences and Antibodies Thereeto
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/724,394A
FILING DATE: 01-OCT-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
```

```
NAME: Fitts, Renee A.
REGISTRATION NUMBER: 35,136
REFERENCE/DOCKET NUMBER: 017957-000100
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-576-0200
TELEFAX: 415-576-0300
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 246240 base pairs
TYPE: nucleic acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..246240
OTHER INFORMATION: /note= "HLA-H.CONFIG"
US-08-724-394A-22
```

```
Query Match          5.7%  Score 31:  DB 2:  Length 246240:
Best Local Similarity 51.0%  Pred. No. 19:
Matches 73:  Conservative 0;  Mismatches 70;  Indels 0;  Gaps 0;
```

```
OY 16 acacattgacgagtcacgcttcctcgtatgtgagtgagtgacgctgcacacg 75
    ||||| ||| ||| ||| ||||| ||| ||||| ||| ||| ||| ||| |||
DB 75473 ACACACTTGCACTCAACAACTTCAGATTATTTCCGCTGCTGACTCACTCCACATTATTC 75532
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 76 tatggcaattctgctgatactcaaccctatgtgtgtgcacaaatgtgcaagtaagcgc 135
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 75533 AGAGCCACTTCCTTTATGGAATATTATGCTTGTCGCAAAAGTATTGCGGCTTTGCC 75592
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 136 aagcaaatcaacgyltaagaacac 158
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 75593 ATTAAAGTAAAGCAAAAACACAC 75615
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 13
US-08-450-351-1
Sequence 1, Application US/08450351
Patent No. 5981213
GENERAL INFORMATION:
APPLICANT: Hansen, Eric J.
APPLICANT: Helminen, Merja E.
APPLICANT: Maciver, Isobel
TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING TO
TITLE OF INVENTION: USEFUL ANTIGENS OF MORAXELLA CATARRHALIS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/450,351
FILING DATE: 25-MAY-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Highlander, Steven L.
REGISTRATION NUMBER: 37,642
REFERENCE/DOCKET NUMBER: AMCY:019
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
```



```

; FILING DATE: December 23, 1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Gravelle, Michelle
; REGISTRATION NUMBER: 40,261
; REFERENCE/DOCKET NUMBER: 6580-81
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 364-7311
; TELEFAX: (416) 361-1398
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3762 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Actinobacillus pleuropneumoniae
; US-08-772-270A-3

```

```

Query Match          5.7%; Score 30.8; DB 3; Length 3762;
Best Local Similarity 49.4%; Pred. No. 3.5;
Matches 80; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 87 tgcctgatgctcaacctatgcttggtgcgaatltggtcaagtagacgcagcaaatcaaa 146
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2100 TGCTATTACGCAACAGCGTGGGATGTCATATCGGTGAACCTTGCGGCATTTACTCGCAA 2159
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 147 cggtaagaacacgcgttatggtattatgcaggttataacttggaccacaatttggcgt 206
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2160 AGGTTCTGATACGAAAGCGGTAAAGCTTACGTTGATTTCTTGAAGAGGAAAGAACTTTT 2219
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 207 agaaccgaaattgctggtcagacgcacaaagaatttaatgc 248
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2220 AGAGAAAGAACCGGATCGTTTGTATATAAAGTCTTTGATCC 2261
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Search completed: March 5, 2001, 06:21:36
 Job time: 17610 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compgen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 4, 2001, 21:46:48 Search time 914.71 Seconds
(without alignments)
4159.860 Million cell updates/sec

Title: US-09-164-714-6
Perfect score: 543
Sequence: 1 atgaataacttaaaaacact.....gcgtccatttgcttttaa 543

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues

Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_estba:*
42: em_estfun1:*
43: em_esthum1:*

44: em_esthum2:*
45: em_esthum3:*
46: em_esthum4:*
47: em_esthum5:*
48: em_esthum6:*
49: em_esthum7:*
50: em_esthum8:*
51: em_esthum9:*
52: em_esthum10:*
53: em_esthum11:*
54: em_esthum12:*
55: em_esthum13:*
56: em_esthum14:*
57: em_esthum15:*
58: em_esthum16:*
59: em_esthum17:*
60: em_esthum18:*
61: em_esthum19:*
62: em_esthum20:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estlov1:*
68: em_estlov2:*
69: em_estp11:*
70: em_estp12:*
71: em_estp13:*
72: em_estp14:*
73: em_estp15:*
74: em_estro1:*
75: em_estro2:*
76: em_estro3:*
77: em_estro4:*
78: em_estro5:*
79: em_estro6:*
80: em_estro7:*
81: em_estro8:*
82: em_estro9:*
83: em_estro10:*
84: em_estro11:*
85: em_estro12:*
86: em_estro13:*
87: gb_estl1:*
88: gb_estl2:*
89: gb_estl3:*
90: gb_estl4:*
91: gb_estl5:*
92: gb_estl6:*
93: gb_estl7:*
94: gb_estl8:*
95: gb_estl9:*
96: gb_estl50:*
97: gb_estl51:*
98: gb_estl52:*
99: gb_estl53:*
100: gb_estl54:*
101: gb_estl55:*
102: gb_estl56:*
103: gb_estl57:*
104: gb_estl67:*
105: gb_estl68:*
106: gb_estl69:*
107: gb_estl70:*
108: gb_estl71:*
109: gb_estl72:*
110: gb_estl73:*
111: gb_estl74:*
112: em_esthum21:*
113: em_esthum22:*
114: em_esthum23:*
115: em_estom1:*
116: em_estom2:*

117: em_estp16:*
 118: em_estp17:*
 119: em_estp18:*
 120: em_estp19:*
 121: em_estp20:*
 122: em_estp21:*
 123: em_estp22:*
 124: em_estp23:*
 125: em_estp24:*
 126: em_estp25:*
 127: em_estp26:*
 128: em_estp27:*
 129: em_estp28:*
 130: em_estp29:*
 131: em_estp30:*
 132: em_estp31:*
 133: em_estp32:*
 134: em_estp33:*
 135: em_estp34:*
 136: em_estp35:*
 137: em_estp36:*
 138: em_estp37:*
 139: em_estp38:*
 140: em_estp39:*
 141: em_estp40:*
 142: em_estp41:*
 143: em_estp42:*
 144: em_estp43:*
 145: em_estp44:*
 146: em_estp45:*
 147: em_estp46:*
 148: em_estp47:*
 149: em_estp48:*
 150: em_estp49:*
 151: em_estp50:*
 152: em_estp51:*
 153: em_estp52:*
 154: em_estp53:*
 155: em_estp54:*
 156: em_estp55:*
 157: em_estp56:*
 158: em_estp57:*
 159: em_estp58:*
 160: em_estp59:*
 161: em_estp60:*
 162: em_estp61:*
 163: em_estp62:*
 164: em_estp63:*
 165: em_estp64:*
 166: em_estp65:*
 167: em_estp66:*
 168: em_estp67:*
 169: em_estp68:*
 170: em_estp69:*
 171: em_estp70:*
 172: em_estp71:*
 173: em_estp72:*
 174: em_estp73:*
 175: em_estp74:*
 176: em_estp75:*
 177: em_estp76:*
 178: em_estp77:*
 179: em_estp78:*
 180: em_estp79:*
 181: em_estp80:*
 182: em_estp81:*
 183: em_estp82:*
 184: em_estp83:*
 185: em_estp84:*
 186: em_estp85:*
 187: em_estp86:*
 188: em_estp87:*
 189: em_estp88:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42	7.7	274	110	BE612078
2	39.4	7.3	1101	190	CNS0182P
3	38.6	7.1	344	89	CNS34656
4	38.6	7.1	519	37	AV553491
5	38.6	7.1	1101	190	CNS00EAX
6	37.8	7.0	748	38	AV715018
7	37	6.8	740	38	AV679268
8	36.6	6.7	587	151	AO256569
9	36.6	6.7	742	38	AV674403
10	36.6	6.7	479	171	AO826019
11	36.4	6.7	481	29	AO83510
12	36.4	6.7	1101	190	CNS008P6
13	36.4	6.6	446	94	AW572255
14	36	6.6	268	92	AW597222
15	35.8	6.6	410	95	AW781475
16	35.8	6.6	426	94	AW760255
17	35.8	6.6	435	94	AW733561
18	35.8	6.6	454	94	AW757253
19	35.8	6.6	460	16	AW138233
20	35.8	6.6	514	165	AO784640
21	35.8	6.6	513	37	AV535288
22	35.6	6.6	549	94	AW754881
23	35.6	6.6	614	191	CNS01WJ0
24	35.6	6.5	392	136	BE839457
25	35.4	6.5	410	38	AV656509
26	35.4	6.5	591	94	AW755155
27	35.4	6.5	624	97	AW981429
28	35.4	6.5	275	129	BB380142
29	35.2	6.5	807	190	CNS00ATD
30	35.2	6.5	611	177	AZ225089
31	35	6.4	1000	190	CNS00COQ
32	35	6.4	407	87	AW201163
33	34.8	6.4	556	150	AO176788
34	34.8	6.4	1086	190	CNS00YXK
35	34.6	6.4	353	147	Z46831
36	34.6	6.4	598	25	AT794544
37	34.6	6.4	801	157	AO449605
38	34.6	6.4	385	93	AW620247
39	34.4	6.3	402	149	AO124099
40	34.4	6.3	428	94	AW761487
41	34.4	6.3	430	94	AW757403
42	34.4	6.3	439	92	AW568592
43	34.4	6.3	442	87	AW202121
44	34.4	6.3	451	94	AW756981
45	34.4	6.3	451	94	AW756981

ALIGNMENTS

RESULT 1
 BE612078
 LOCUS BE612078 274 bp mRNA
 DEFINITION sro3e08.y1 Gm-c1049 glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1049-1599 5' similar to FR:080504 080504 F16h22.14 PROTEIN. ;,
 mRNA sequence.
 BE612078
 ACCESSION BE612078.1 GI:9903110
 VERSION
 KEYWORDS EST.

Accession	Version	KeyWords	Source	Organism
AM334656.1	GI:6831117	sequence.		
AM334656				
AM334656.1	GI:6831117	EST.		
Pneumocystis carinii f. sp. carinii.				
Pneumocystis carinii f. sp. carinii.				
Eukaryota; Fungi; Fungi Incertae sedis; Pneumocystidaceae;				
Pneumocystis.				
1 (bases 1 to 344)				
Smullen,A.G., Arnold,J., Weise,M., Wunderlich,J., Staben,C., Edman				
,J.C., Kovacs,J. and Cushion,M.				
Expressed sequence tags from Pneumocystis carinii				
Unpublished (2000)				
Contact: Staben C				
School of Biological Sciences				
University of Kentucky				
101 Morgan Building, University of Kentucky, Lexington, KY				
40506-0225, USA				
Tel: 606 257 2161				
Fax: 606 257 1717				
Email: staben@pop.uky.edu.				
Location/Qualifiers				
1. .344				
/db_xref="Pneumocystis carinii f. sp. carinii"				
/db_xref="taxon:38081"				
/clone_lib="AGS-1"				
/lab_host="E. coli"				
/note="Vector: Lambda ZAP II; Site_1: EcoRI; Site_2: XhoI				
P. carinii organisms (3x10e9) from a single rat (99-1-6,				
sacrificed on 3/17/99) at Cincinnati VA facilities.				
Trisol extracted RNA. Oligo dT priming, standard				
conditions described by vendor, Stratagene. Further				
details see www.uky.edu/Project/Pneumocystis/"				
BASE COUNT	99 a	43 c	50 g	152 t
ORIGIN				
Query Match	7.1%	Score 38.6:	DB 89:	Length 344:
Best Local Similarity	47.3%:	Prod. No. 0.49:		
Matches 116:	Conservative 0:	Mismatches 129:	Indels 0:	Gaps 0
QY 127	gtagagcccaagcaaatcaacgtagaagaccgcttagtattatgtagtataac	186		
DB 253	GAAAACGCTTAATATTATATAAAGAAATCTAATCTTATGAACACGCGGTATGTATA	194		
QY 187	tttagcccaaatlttggcgtagaacccgaatttgttgctcgaagccaaagaattaat	246		
DB 193	ATTGAACCAAAATCTTTGTTTTCGGGAACCGAACCTCATATTGCAAAAAATTTAAT	134		
QY 247	gcagcgctgagtcctcgtaaaagtgatgtagcttlttggtgcttagtgcacatctgc	306		
DB 133	GTCGTATATCATGTGCGCAAAAGAGTAATAAAATTCATTTTATCTCAGCATCTTTAGAC	74		
QY 307	tataacttcataabacccattttagccaaagggcaaattaggcattgctaagactaa	366		
DB 73	GAAAATTCCTTCAAAAAACCAAGTATTATCAATGGAAGAGAGATTTCCTCAAAAAA	14		
QY 367	gtaga 371			
DB 13	AAAAA 9			
RESULT 4				
AV553491	519 bp	mRNA	EST	06-SEP-2000
LOCUS				
DEFINITION	AV553491	Arabidopsis thaliana roots Columbia	Arabidopsis thaliana	
ACCESSION	AV553491			
VERSION	AV553491.1	GI:8724904		
KEYWORDS	EST.			
SOURCE	thale cress.			
ORGANISM	Arabidopsis thaliana			
	Eukaryote; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;			

REFERENCE	Magnoliophyta: eudicotyledons: core eudicots: Rosidae: eurosid 11
AUTHORS	Brassicales: Brassicaceae: Arabidopsis.
TITLE	1 (bases 1 to 519) Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S. A large scale analysis of cDNA in Arabidopsis thaliana: Generation of 12,028 non-redundant expressed sequence tags from normalized size-selected cDNA libraries DNA Res. 7, 175-180 (2000)
JOURNAL	20363093
MEDLINE	Contact: Erika Asamizu The First Laboratory for Plant Gene Research Kazusa DNA Research Institute Yama 1532-3, Kisarazu, Chiba 292-0812, Japan Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/ Location/Qualifiers
FEATURES	1..519 /organism="Arabidopsis thaliana" /strain="Columbia" /db_xref="taxon:3702" /clone="R263g09R" /clone_lib="Arabidopsis thaliana roots Columbia" /tissue_type="roots" /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT	144 a 90 c 122 g 163 t
ORIGIN	
Query Match	7.1%; Score 38.6; DB 37; Length 519;
Best Local Similarity	51.4%; Pred. No. 0.55;
Matches	89; Conservative 0; Mismatches 84; Indels 0; Gaps 0
OY	122 gtcacgtacgcacgaacatcaacacgctaaacacccgtataglatlalcgaagtt 181
Db	336 GTTTAGTTGTGACCACTTGTCGAAGAAGACTAGAGACTGATGACTAATGAAGCTT 385
OY	182 ataacttcgacaaaatttgcgcagaaaccgaatttgcttcagacgcacaaagt 241
Db	366 CTTCTCTTGAGGATTAAGATTACTTATGATTTGATTTCTAGAAATTTAGATGCAAGAAT 445
OY	242 ttaatcgcagcgtcgagtcctgtaaaaagctgaatgaaagctcttgcgtcatal 294
Db	446 CTCACGGTGAACCAATGCCGACGATATTGATCTGGAACCTTCTAAGATGAT 498
RESULT	5
CNS00EAX/C	CNS00EAX 1101 bp DNA GSS 04-JUN-1999
LOCUS	Drosophila melanogaster genome survey sequence Ttr3 end of BAC #
DEFINITION	BACR2BD24 of RPc1-98 library from Drosophila melanogaster (fruit
ACCESSION	fly), genomic survey sequence.
VERSION	AL068861
KEYWORDS	AL068861.1 GI:4949105
SOURCE	GSS.
ORGANISM	fruit fly.
REFERENCE	Drosophila melanogaster
AUTHORS	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
TITLE	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
JOURNAL	Muscomorpha; Ephyroidae; Drosophilidae; Drosophila.
COMMENT	1 (bases 1 to 1101) Genoscope. Direct Submission Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr) Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazuo Oseegawa and Aaron Mammosser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo.

QY 236 aagaattlaatgcaagcgtagctcctgtaaaaggtagatgtagaglcctlllgtagcttat 294
||| | ||| |
Db 150 GAGAACTAAAGGCTGGAACCAAGGTACTATTCACATGACATGAATCCTTAAAGAGTGGAT 208
||| | ||| | ||| | ||| |

Search completed: March 5, 2001, 03:19:25
Job time: 19957 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 14:36:12 ; Search time 26.49 Seconds
(without alignments)
232.347 Million cell updates/sec

Title: US-09-164-714-7
Perfect score: 912
Sequence: 1 MKTLKTLAVASASLLAMSA.....SYNYLSEDAVAISGLHLP 180

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues
Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_36:*

- 1: /SIDSI/gcgdata/geneseq/geneseqp/AA1980.DAT:*
- 2: /SIDSI/gcgdata/geneseq/geneseqp/AA1981.DAT:*
- 3: /SIDSI/gcgdata/geneseq/geneseqp/AA1982.DAT:*
- 4: /SIDSI/gcgdata/geneseq/geneseqp/AA1983.DAT:*
- 5: /SIDSI/gcgdata/geneseq/geneseqp/AA1984.DAT:*
- 6: /SIDSI/gcgdata/geneseq/geneseqp/AA1985.DAT:*
- 7: /SIDSI/gcgdata/geneseq/geneseqp/AA1986.DAT:*
- 8: /SIDSI/gcgdata/geneseq/geneseqp/AA1987.DAT:*
- 9: /SIDSI/gcgdata/geneseq/geneseqp/AA1988.DAT:*
- 10: /SIDSI/gcgdata/geneseq/geneseqp/AA1989.DAT:*
- 11: /SIDSI/gcgdata/geneseq/geneseqp/AA1990.DAT:*
- 12: /SIDSI/gcgdata/geneseq/geneseqp/AA1991.DAT:*
- 13: /SIDSI/gcgdata/geneseq/geneseqp/AA1992.DAT:*
- 14: /SIDSI/gcgdata/geneseq/geneseqp/AA1993.DAT:*
- 15: /SIDSI/gcgdata/geneseq/geneseqp/AA1994.DAT:*
- 16: /SIDSI/gcgdata/geneseq/geneseqp/AA1995.DAT:*
- 17: /SIDSI/gcgdata/geneseq/geneseqp/AA1996.DAT:*
- 18: /SIDSI/gcgdata/geneseq/geneseqp/AA1997.DAT:*
- 19: /SIDSI/gcgdata/geneseq/geneseqp/AA1998.DAT:*
- 20: /SIDSI/gcgdata/geneseq/geneseqp/AA1999.DAT:*
- 21: /SIDSI/gcgdata/geneseq/geneseqp/AA2000.DAT:*

prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	912	100.0	180	21	Y84612
2	209	22.9	40	21	Y84611
3	152	16.7	31	21	Y56863
4	115	12.6	188	13	R26415
5	112.5	12.3	361	11	R05999
6	112.5	12.3	361	11	R66877
7	112	12.3	359	16	R66294
8	112	12.3	362	10	P90098
9	108	11.8	188	13	R25674
10	108	11.8	188	16	R70744
11	108	11.8	188	18	W18380
12	106.5	11.7	162	17	R96207

13	106.5	11.7	338	16	R85450
14	106.5	11.7	568	17	R96210
15	100	11.0	288	18	W27720
16	100	11.0	1287	16	R79944
17	97	10.6	211	18	W55324
18	97	10.6	211	18	W20755
19	95.5	10.5	1290	19	W98269
20	95	10.4	1288	18	W55547
21	95	10.4	1288	18	W55685
22	94.5	10.4	390	18	W21678
23	93.5	10.3	287	18	W53462
24	91	10.0	1296	14	R41198
25	89.5	9.8	767	19	W46272
26	89.5	9.8	1267	11	R04232
27	88.5	9.7	174	17	W04891
28	87.5	9.6	443	20	W94460
29	86.5	9.5	363	16	R66878
30	85.5	9.4	711	19	W73025
31	85.5	9.4	758	19	W46270
32	85	9.3	621	18	W55645
33	85	9.3	621	20	W17209
34	85	9.3	637	18	W55739
35	85	9.3	637	20	W17208
36	85	9.3	691	19	W73032
37	85	9.3	691	19	W73032
38	84.5	9.3	255	18	W55477
39	84.5	9.3	255	18	W55201
40	84.5	9.3	255	20	W17205
41	84	9.2	174	17	W04893
42	83	9.1	759	19	W46271
43	83	9.1	759	19	W46273
44	83	9.1	759	21	Y52396
45	83	9.1	759	21	Y52136

ALIGNMENTS

RESULT 1	
ID Y84612	Standard: Protein: 180 AA.
XX Y84612:	
AC	
XX	
25-JUL-2000 (first entry)	
DT	
XX	
DE	The outer membrane protein 21 of strain ATCC49143.
XX	
XX	Outer membrane protein 21; OMP21; strain ATCC49143; bacterial infection;
KW	otitis media; respiratory infection; sinusitis; pneumonia; immunisation.
XX	
OS	Moraxella catarrhalis.
XX	
FH	Key
FT	Misc-difference 71
FT	Location/Qualifiers
XX	
PN	W0200018910-A1.
XX	
PD	06-APR-2000.
XX	
PF	01-OCT-1999; 99WO-US22918.
XX	
PR	01-OCT-1998; 98US-0164714.
XX	
PA	(ANTE-) ANTEX BIOLOGICS INC.
XX	
PI	Tucker K, Tillmann UF;
XX	
DR	WPI: 2000-293149/25.
XX	
DR	N-PSDB: A12591.
XX	
PT	Isolated outer membrane protein from a Moraxella catarrhalis strain

PT used for diagnosis treatment and prevention of disease caused by M.
XX catarrhalis e.g. pneumonia, otitis media and respiratory infections -
XX
XX
XX Claim 2; Fig 4; 108pp; English.

The present sequence represents an outer membrane protein 21 (OMP21) of Moraxella catarrhalis strain ATCC91943. The OMP21 protein has an apparent molecular weight of 16-20 kD as determined by sodium dodecylsulphate polyacrylamide gel electrophoresis (SDS-PAGE). OMP21, its nucleic acids and antibodies can be used in prophylactic and therapeutic compositions for treating a M. catarrhalis bacterial infection, otitis media, respiratory infections, sinusitis and pneumonia. They are useful as reagents for the clinical or medical diagnosis of M. catarrhalis infections and for scientific research on the properties of pathogenicity, virulence and infectivity of M. catarrhalis and host defence mechanisms. The antibodies, particularly those that are cytotoxic may be used in passive immunisation to prevent or attenuate M. catarrhalis infections of animals e.g. humans.

Sequence 180 AA;

Query Match	100.0%;	Score 912;	DB 21;	Length 180;
Best Local Similarity	100.0%;	Pred. No. 2,66-82;		
Matches 180; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

QY 1 MKTLATLLAVSASSLLAMSNANAISYGNADAPQPYVGAKIQVDAKQINGKNTAYGIYAC 60
Db 1 mktlktllavassllamsanaatsygnadapqpyvaklqvddakqlnknrtayvqiyaq 60			

QY 61 YNFDQNEGVEAEFVGSDAKEFNAGVSPVKGDVKSFGAYGTYRYNFINTPPYAKGKLGIAK 120

Db 61 ynfddnfgveaefvgdsakefnagvspvkqdvksfgaygtyrnfintpfyakgkigiak 120

QY 121 TKVDVTSRNA^{TT}YSNKS^{SD}KTSLAGVG^{VG}E^{FK}PLANVG^{VS}SYN^{TL}ISEDANATSLGAHLAF 180

Db 121 tkvdtsrnat^{ty}snks^{dk}tslagvgv^gfkplanvgv^{es}syn^{tl}sedana^{ti}slgahlaf 180

RESULT	2
Y84611	
ID	Y84611 standard; peptide; 40 AA.

AC Y84611;

DT 25-JUL-2000 (first entry)

DE N-terminal peptide of outer membrane protein 21 of strain ATCC49143.

Outer membrane protein 21; OMP21; strain ATCC49143; bacterial infection; otitis media; respiratory infection; sinusitis; pneumonia; immunisation.

05 Moraxella catarrhalis.

PN WO200018910-A1.

PD 06-APR-2000.

PF 01-OCT-1999; 99WO-US22918.

PR 01-OCT-1998; 98US-0164714.

PA (ANTE-) ANTTEX BIOLOGICS INC.

PI Tucker K, Tillmann UF;

DR WPI; 2000-293149/25.

PT Isolated outer membrane protein from a *Moraxella catarrhalis* strain

PT catarhalis e.g. pneumonia, otitis media and respiratory infections -

PS Claim 2; Page 63; 108pp; English.

XX The present sequence represents the N-terminal peptide of a mature
CC outer membrane protein 21 (OMP21) of *Moraxella catarrhalis* strain
CC ATCC49143. The specification describes an OMP21 from a *Moraxella*
CC *catarrhalis* strain, which has an apparent molecular weight of 16-20 kD
CC as determined by sodium dodecylsulfate polyacrylamide gel electrophoresis
CC (SDS-PAGE). OMP21, its nucleic acids and antibodies can be used in
CC prophylactic and therapeutic compositions for treating a *M. catarrhalis*
CC bacterial infection, otitis media, respiratory infections, sinusitis
CC and pneumonia. They are useful as reagents for the clinical or medical
CC diagnosis of *M. catarrhalis* infections and for scientific research on
CC the properties of pathogenicity, virulence and infectivity of
CC *M. catarrhalis* and host defence mechanisms. The antibodies, particularly
CC those that are cytotoxic may be used in passive immunisation to prevent
CC or attenuate *M. catarrhalis* infections of animals e.g. humans.
XX

SQ Sequence 40 AA;

Query Match	22.9%	Score 209;	DB 21;	Length 40;
Best Local Similarity	100.0%	Pred. No. 4.4e-14;		
Matches 40;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

```

QY      23 AISTGNSADAPYVGAkIGQYDAKQINGKNTAYGIAGYN 62Z
      |||||
Db      1 aistgnsadapyyvakigqydaKqngkntaygiagyn 40C

```

RESULT	3
Y56863	
ID	Y56863 standard; peptide; 31 AA

DT 10-APR-2000 (first entry)

DE B. catarrhalis 20 kD antigen N-terminal fragment.

KW Branhamella catarrhalis; Moraxella catarrhalis; antigen; vaccine;

XX	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
XX	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100

XX

XX

XX

XX

XX

XX

XX

XX

PT and in vaccines -

PS Claim 5; Page 22; 32pp; English.

The invention relates to novel *Branhamella catarrhalis* (Moraxella catarrhalis) antigens having an apparent molecular weight of about 14-17 kDa (as determined by SDS-PAGE). The antigens can be used to prepare vaccines and immunogenic compositions for the treatment and prophylaxis of *Branhamella catarrhalis* infections, respiratory tract infections, and otitis media. Antibodies against the antigens can be used for diagnosis and purification of the antigens. Sequences 15683-871 represent peptide fragments from the *B. catarrhalis* antigens of the invention.

Sequence	31 AA;
5Q	

```
Query Match          16.7%; Score 152; DB 21; Length 31;
Best Local Similarity 96.8%; Pred. No. 1.3e-08;
Matches 30; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 23 A1SYGNSADAPYVAKIGOVDAKQINGKNT 53
    |||
Db 1 a1sygnasadpyvgaklgyvdakqimnkt 31

RESULT 4
R26415
ID R26415 standard; Protein: 188 AA.
XX
AC R26415;
XX
DT 17-MAR-1993 (first entry)
XX
DE pagC/AP fusion protein.
XX
KW Salmomella; pagC; alkaline phosphatase; AP; transposon; TnpHoA;
KW CS119; phoP; regulatory region regulated gene; virulence.
XX
OS Salmomella typhimurium.
XX
XX
FH Key Location/Qualifiers
FT Peptide 1..23
FT Protein /Label= Signal_peptide 24..188
FT Region /Label= Mature_protein 1..33
FT Region /Label= pagC 34..188
FT Region /Label= AP
XX
PN WO9217785-A.
XX
PD 15-OCT-1992.
XX
PF 30-MAR-1992; 92MO-US02591.
XX
PR 29-MAR-1991; 91US-0678409.
XX
PA (GEHO ) GEN HOSPITAL CORP.
XX
PI Miller SI;
XX
DR WPI: 1992-366389/44.
DR N-PSDB; Q27739.
XX
PT Detection of Salmomella using monoclonal antibodies - against pho
PT P regulated gene product useful for diagnosing e.g. typhoid
PT fever, enteric fever(s) acute gastroenteritis etc.
XX
PS Disclosure: Fig 2; 39pp; English.
XX
CC This sequence represents a fusion between the Salmomella pagC protein
CC and the alkaline phosphatase (AP) protein. The AP gene was inserted
CC into the pagC gene on the transposon TnpHoA. Insertion of this
CC transposon into pagC mutant strain CS119 caused interruption of an
CC 18kD protein to form the new protein species of approx. 45 kD given.
CC pagC is a phoP regulatory region regulated gene and a mutation at the
CC pagC locus confers a virulence defect on Salmomella typhimurium.
XX
SQ Sequence 188 AA;

Query Match          12.6%; Score 115; DB 13; Length 188;
Best Local Similarity 27.1%; Pred. No. 0.00058;
Matches 55; Conservative 28; Mismatches 82; Indels 38; Gaps 11;
OY 1 MTKTLKLLVASSSLAMSNAA1SYGNSADAPYVGAIGOVDAKQINGKNTAYIGAG 60
    |||
```

```
Db 1 mknlislclvltslvlnvagaqlnaifsvgyapyaqskv--qdflnirgvnvyk-----r 54
OY 61 YNFDONFGVEAEFVSGSDA-----KEFNAGVSPV-----KGDVKSFGAYCT-----YRY 103
    |||
Db 55 yedds-----pvsflsrlslyldrsgasgsvepeglyhdkfewk-----ygsilmvgyrfl 106
OY 104 NFINTPEFYAKGKGLAK-TKVDVTSRNATYTSNK--SDKTSLAGVGVGFKPLANGVEA 160
    :|||
Db 107 s-dfslalyalagvgvkvkatefkstqgdsfsnklsrktgfawgagvqgmplentlvdv 165

OY 161 SY---NYISEDANATSLGAHLAF 180
Db 166 gyegsnisstklnglnvgvgyrt 188

RESULT 5
R05999
ID R05999 standard; Protein: 361 AA.
XX
AC R05999;
XX
DT 04-DEC-1990 (first entry)
XX
DE P2 gene product of Haemophilus strain 1H.
XX
KW P2 gene; strain 1H; influenza; vaccine; ds.
XX
OS Haemophilus influenzae.
XX
PN EP378929-A.
XX
PD 25-JUL-1990.
XX
PF 22-DEC-1989; 89EP-0313573.
XX
PR 01-FEB-1989; 89GB-0002178.
PR 23-DEC-1988; 88GB-0030124.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Munson RS, Tolan RW, Chong P, Fahim R, McVerry P, Klein M;
XX
DR WPI: 1990-225607/30.
DR N-PSDB; Q05372.
XX
PT Gene coding for protein P2 of Haemophilus influenzae type-B -
PT used for developing vaccines for protection against disease
PT caused by the organism.
XX
PS Claim 1; Page 8; 15pp; English.
XX
CC Peptide may be used as a vaccine to the disease caused by H1
CC type b, as carrier for conjugation to oligosaccharide derived
CC from Haemophilus. Protein may be used with other haptens as
CC T-cell dependant antigen and carrier.
XX
SQ Sequence 361 AA;

Query Match          12.3%; Score 112.5; DB 11; Length 361;
Best Local Similarity 28.1%; Pred. No. 0.0024;
Matches 52; Conservative 25; Mismatches 65; Indels 43; Gaps 11;
OY 5 KTLVAVSASSSLAMSNAA1SYGNSADAPYVGAIGOV-----DAKQING-----K 51
    |||
Db 3 ktlavlvgaafaasanaavvyneg-tnvelggrlsllaegsnstvdnqkqgslrng 61
OY 52 NTAVGIVAGYNFDONF-----GVAEFV-----GSDAKFENAGVSPVKSGFAGYGY 101
    :|||
Db 62 gsrflhkachnfgdgyagyletrfvtkasengsd-nf-----gdits-----ky 106
OY 102 RYNEFITPPYAKGLGIATKVD-VTSRNATYTS--NKSOKTSLAGG-VGVGFKPLANGV 157
    |||
```



```

Query Match      11.8% Score 108: DB 16: Length 188:
Best Local Similarity 24.2%: Pred. No. 0.0028:
Matches 50: Conservative 29: Mismatches 82: Indels 46: Gaps 8:

QY      1 MKTLKTLTLLAVSASSLLAASANAATISYGSNADAQPYVGAKIGGVDAKQIINGKWTATAGTAYAG 60
      || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1 mkhllstclvtltstlylvnvnvagadlnafsvayaryagskv--qdklnlrtyvnxkY----- 53

QY      61 YNFDPNFCVEAEFVQSDDAKEFNAGSPPVKCGVKSFGA-----YGT----- 100
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      54 -----ryeddpvstfsslsylygdrqasgsrtepegihyhdckfevkvyslmvyp 102

QY      101 -RYRNFINTPPYANGKGLGIAF-TKVDVTSRNAATYYSNK--SDKTSLAGVGVGFGKPLNAV 156
      || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      103 ayrlls-dnfslyalagvytkatfkshstqgdsfsnkslstrktgfwagayvqmplenl 161

QY      157 GVEASY---NYLSEDAANAISLGAIHLAF 180
      || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      162 vvdvgyegsnisslkingfnvgygyrf 188

RESULT 11
W18380
ID      W18380 standard; Protein; 188 AA.
EX

```

Query Match	11.8%	Score 108;	DB 18;	Length 188;
Best Local Similarity	24.2%;	Pred. No.	0.0028;	
Matches	50;	Conservative	29;	Mismatches 82;
			Indels	46; Gaps
OY	1 MKTKTLAVASAILLAMSANAATSYGSNADQPPVGAIGTGVDAKQINGKRTAYGIAG	60		
Dd	1 mknlslstlvtlsylvnvvaqadnalsfvgaryaqskv--gdfknlrgrnvkv-----	53		
OY	61 YNFQDNFVGAEAFEGVSADKEFNAGSVPRVKGDVKSFGA-----YGT-----	100		
Dd	54 -----rlyedspvfstsslsylygdrgaaqsvepegiyhndkfevkyslmvyp	102		
OY	101 -YRNFINTPFYAKCKLGIAK-TKDVTISRNTTYSNK--SPKTSIAGVGCVGFPLANV	156		
Dd	103 ayrll-dnfsljalayagvlvatrfefhehtqdqdsfsnklsrtktgfawgaaygmpleni	161		
OY	157 GVEASY---NYLSEDNAISLGAAHLAF	180		
Dd	162 vdvdygegsnlstekingfnvgvygrf	188		
RESULT	12			
R96207				
ID	R96207 standard; Protein; 162 AA.			
XX				

```

AC R96207;
XX
XX 24-AUG-1996 (first entry)
DE Attachment-Invasion-locus protein.
XX
XX Attachment-Invasion-locus protein; drug delivery;
KM gastrointestinal membrane; transcytosis; bioavailability;
KM fusion protein; enterocyte; Peyer's patch M-cell.
XX
OS Versinia enterocolitica strain 8081c.
XX
XX Key Location/Qualifiers
FH Peptide 19..34
FT /note= "Peptide from extracellular loop-1"
FT Peptide 58..76
FT /note= "Peptide from extracellular loop-2"
FT Peptide 102..119
FT /note= "Peptide from extracellular loop-3"
FT Peptide 145..152
FT /note= "Peptide from extracellular loop-4"
XX
XX W09613250-A1.
XX
XX 09-MAY-1996.
XX
XX 20-OCT-1995; 95WO-US13749.
XX
XX 27-OCT-1994; 94US-0331393.
XX
XX (AMGE-) AMGEN INC.
XX
XX Habberfield AD, Jensen-Pippo K;
XX
XX WPI: 1996-251447/25.
XX
XX N-PSDB: T27536.
XX
XX
XX Therapeutic delivery system utilising bacterial invasion protein - is
XX not readily degraded in the gut, enhances systemic bio-availability
XX of therapeutic agents
XX
XX Example 1; Fig 2; 110pp; English.
XX
XX The sequence represents a bacterial attachment-invasion-locus (AIL)
XX protein, which may be complexed with a therapeutic agent to
XX transport the agent across the gastrointestinal membrane barrier by
XX transcytosis to increase bioavailability 5- to 100-fold. The agent
XX and bacterial protein may be linked via a degradable peptide
XX sequence as a fusion protein. The receptor binding region of the
XX AIL protein involves all or some of the regions from the 4
XX extracellular loops. This region retains the binding affinity of
XX the protein, and may be used alone or as part of a fusion protein
XX for drug delivery. The protein may be fused with maltose binding
XX protein (R96208) to form a fusion protein (R96209) which may be
XX purified easily by amylose affinity chromatography. The delivery
XX system allows improved transport across enterocytes and Peyer's
XX patch M-cells. The system is not prone to degradation in the gut or
XX early release of biologically active material, and eliminates the
XX need for parenteral administration.
XX
XX Sequence 162 AA;

```

```

Query Match 11.7%; Score 106.5; DB 17; Length 162;
Best Local Similarity 30.1%; Pred. No. 0.0033;
Matches 55; Conservative 19; Mismatches 72; Indels 37; Gaps 10;

```

```

OY 10 VSASSLLAMSAANAISY---GNSADAOPYVCAKIGOVDAKQINGKNTATYAGVNFQ 65
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 5 vyaaesaeisigyagshvkengytlndp-----kgfllky-----ryeldd 46
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 66 NPGVGEAFVGS-DAKEFNAGVSPV-KGDVKSFGAYGTYRYNF-INTPFYAKGLGIATKR 122
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

```

Db 47 nwgvgisfaytlbggydflygsnkgfhgwdvdyv--tmgpsfirneyalygl1gaahg 104
OY 123 VDVTSRNATTY---SNKSDSTSLAGVGGEKPLANYGVEASYNYLSEDAANAIS---LGAH 177
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 105 v-----kasvdeslsasaktsmayagvgfnlpnlvidasjyskldskvgtm1gag 159
OY 178 LAF 180
|
Db 160 yrf 162

```

```

Query Match 11.7%; Score 106.5; DB 16; Length 338;
Best Local Similarity 23.6%; Pred. No. 0.0086;
Matches 50; Conservative 18; Mismatches 64; Indels 63; Gaps 11;

```

```

OY 35 YVGAKIGQ---VAKQING-----KNT-AYGIYAGYNF--DONFGVEAE 72
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 8 yayvkagqggsfhdginnngakidsldltlgygyrntltfygfygqqlingdnfglaae 67
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 73 FVGSID-----AKEFNAGVS-PVKGDVKSFGAYGTYRYNF-INTPFYAKGLG 117
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```


THIS PAGE BLANK (USPTO)

QY 99 -GTYRYNFINPFYAKGLGIATKTV--VTSRNATYYSNKSOKTSLAG--VGCGFK 151
Db 117 KGSYE--VLGGLDYYGKAGVALVSDKYEDANGTRDHKKGRHTARASGLFVAGAEYA 173
OY 152 PLANGVEASYNYS 166
Db 174 VLPELAVRLEYOWLT 188

RESULT 2
US-08-467-722A-2

; Sequence 2, Application US/08467722A
; Patent No. 6030626
; GENERAL INFORMATION:
; APPLICANT: Kolattukudy, P. E.
; TITLE OF INVENTION: Otitis Media Vaccine
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESS: Calfee, Halter and Griswold
; STREET: Suite 1800 800 Superior Avenue
; CITY: Cleveland
; STATE: Ohio
; COUNTRY: U.S.A.
; ZIP: 44114-2688
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PC-DOS/MS-DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/467,722A
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Golfick, Mary E.
; REGISTRATION NUMBER: 34,829
; REFERENCE/DOCKET NUMBER: 22727/00102
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (216) 622-8458
; TELEFAX: (216) 241-0816
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 359 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-467-722A-2

Query Match 12.4%; Score 113; DB 3; Length 359;
Best Local Similarity 28.2%; Pred. No. 0.00014;
Matches 55; Conservative 18; Mismatches 80; Indels 42; Gaps 11;

OY 5 KTLAVASASSILAMSANAIAISYNSADQPIYVAKIGQ---VDAKQING----- 50
Db 3 KTAIALVYAGLAASVAQAAPQENTF---YAGYKAGGSGSHDGINNNGAIKKGLSSSNY 58
OY 51 --KNT-AYGIYAGYNF--DONFGVEAEFYVGS--AKEFNAGVSPYKGVKSGFAGY-- 98
Db 59 GYRNRTFPGYGVFGQILINDNPGLAELGYDDFGRAKLRAGKPKAKH--TNNGAYISL 116
OY 99 -GTYRYNFINPFYAKGLGIATKTV--VTSRNATYYSNKSOKTSLAG--VGCGFK 151
Db 117 KGSYE--VLGGLDYYGKAGVALVSDKYEDANGTRDHKKGRHTARASGLFVAGAEYA 173
OY 152 PLANGVEASYNYS 166
Db 174 VLPELAVRLEYOWLT 188

RESULT 3
US-08-096-181A-8

; Sequence 8, Application US/08096181A
; Patent No. 6153406
; GENERAL INFORMATION:
; APPLICANT: Tai, Joseph Y.
; APPLICANT: Pullen, Jeffrey K.
; APPLICANT: Soper, Thomas S.
; APPLICANT: Liang, Shu-Mei
; TITLE OF INVENTION: A Method For The High Level Expression,
; TITLE OF INVENTION: Purification And Refolding Of The Outer Membrane Protein
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESS: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PC-DOS/MS-DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/096,181A
; FILING DATE: 23-Jul-1993
; CLASSIFICATION: 424
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 361 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-096-181A-8

Query Match 12.3%; Score 112.5; DB 3; Length 361;
Best Local Similarity 28.1%; Pred. No. 0.00016;
Matches 52; Conservative 25; Mismatches 65; Indels 43; Gaps 11;

OY 5 KTLAVASASSILAMSANAIAISYNSADQPIYVAKIGQY-----DAKQING---K 51
Db 3 KTLAALIVGAFPAASANAAYVYNNEG--TNVELGGGLSTIAOSNSTYDNRQOQGLALRNQ 61
OY 52 NTAYGIYAGYNFQDNF---GVEAEFY-----GSDAKEFNAGVSPYKGVKSGFAGTY 101
Db 62 GSRPIKATNHFQDGFYAGQVLETRFYTKASENGSD--NF-----GDITS-----KY 106
OY 102 RYNFINPFYAKGLGIATKTV--VTSRNATYYS--NKSOKTSLAG--VGCGFKPLANVG 157
Db 107 AYVTLGNKAPEVKLRGAKKTLADGITSADKEYGVYLNNSDYIPIISGNTVGTTFGIDSLV 166
OY 158 YEAST 162
Db 167 LGANY 171

RESULT 4
PCT-US94-08326-8

; Sequence 8, Application PC/TUS9408326
; GENERAL INFORMATION:
; APPLICANT: North American Vaccine, Inc.
; APPLICANT: 12103 Indian Creek Court
; APPLICANT: Beltsville, MD 20705
; APPLICANT: Pullen, Jeffrey K.
; APPLICANT: Soper, Thomas S.
; APPLICANT: Liang, Shu-Mei
; TITLE OF INVENTION: A Method For The High Level
; TITLE OF INVENTION: Expression,
; TITLE OF INVENTION: Purification And Refolding Of The Outer Membrane

;; FILING DATE:
;; CLASSIFICATION:
;; INFORMATION FOR SEQ ID NO: 5:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 568 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: unknown
;; TOPOLOGY: unknown
;; MOLECULE TYPE: protein
PCT-US95-13749-5

Query Match 11.7%; Score 106.5; DB 4; Length 568;
Best Local Similarity 30.1%; Pred. No. 0.0014;
Matches 55; Conservative 19; Mismatches 72; Indels 37; Gaps 10;

QY 10 VSASSILMSANNAISY-----GNSADAPYVGAKIGVDAKQINKNRYAGIYAGNFDQ 65
Db 411 YVAASSESSISIGYQSHKENGYTLDNDP-----KGFNLKY-----RYELDD 452
QY 66 NFGVAFVGS-DAKEFNAGVSPV-KGVKSGFAGYTRYRNF-INPPYAKGLGIATKR 122
Db 453 NMGVGSFAYTHQGYDFFGSKNKGCHDVYYSV--TWGSPFRINEYVSLYGLGAAGHK 510
QY 123 VDVTSRNATTY--SNKSDKSLAGSGYGFKPLANVGEASYNILSEDAANIS--LGAH 177
Db 511 V-----KASVEDESISAKSTMAAGAGVQFNPDPFVIDASYSKLDSIKVGTWMLGAG 565
QY 178 LAF 180
Db 566 YRF 568

RESULT 7
US-08-210-394-1
; Sequence 1, Application US/08210394
; Patent No. 5770213
; GENERAL INFORMATION:
; APPLICANT: Zlocnick Dr., Gary W.
; TITLE OF INVENTION: Purified No. 5770213typable Haemophilus
; TITLE OF INVENTION: Influenzae p5 Protein as a Vaccine for No. 5770213typable
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: American Cyanamid Company
; STREET: One Cyanamid Plaza
; CITY: Wayne
; STATE: New Jersey
; COUNTRY: US
; ZIP: 07470-8426
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/210.394
; FILING DATE: 07-MAR-1994
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Harrington, James J
; REFERENCE/DOCKET NUMBER: 32,144
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201/831-3246
; TELEFAX: 201/831-3305
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 338 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-210-394-1

Query Match 11.6%; Score 105.5; DB 1; Length 338;
Best Local Similarity 25.6%; Pred. No. 0.00085;
Matches 50; Conservative 17; Mismatches 65; Indels 63; Gaps 11;

QY 35 YVGAKIGQ---VDAKQING-----KNT-AVGIAGYWF--DQNFGEAE 72
Db 8 YAGVACGCGSFHDGINNGAIKEDSIDLTGYYRRNFTFYGVFGYQILNQDNFGLAE 67
QY 73 FVGSD-----AKEFNAGVS-PYKGVKSGFAGYTRYRNF-INPPYAKGLG 117
Db 68 -LGDNFGFRVYFRAEGTKAKHTNHGAHLSTLKGSTIYLDGIDYV-----GKAG 114
QY 118 IAKTRVD--VTSRNATTYSNKSDKTSLAG--VGVGFKPLANVGEASY-----NY 164
Db 115 VALVMSDYKFEAPRSTDRDXKKGHTARASGLPANGAEAVLPFLAVLLEFVQQLTRVGKY 174
QY 165 LSEDAANISLGAHLA 179
Db 175 RPODKNAPSINPNTA 189

RESULT 8
US-08-486-715-5
; Sequence 5, Application US/08486715
; Patent No. 5674724
; GENERAL INFORMATION:
; APPLICANT: Miller, Samuel I.
; TITLE OF INVENTION: SALMONELLA VIRULENCE GENES
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P. C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486.715
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/090.526
; FILING DATE: 09-JUL-1993
; PRIOR APPLICATION NUMBER:
; APPLICATION DATA:
; FILING DATE: 07-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00786/192005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 188 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-486-715-5

Query Match 11.3%; Score 103; DB 1; Length 188;
Best Local Similarity 27.0%; Pred. No. 0.00071;

Db 55 YEDDSPVSPFSSLSLYLGD-RQASGSVEPEGIHYHDKFEVKKGSLMVGPAVRLS-DNFSL 112
OY 111 YAKKGLGIAT-KYVDVTSRNATTTYSNK--SDKTSLAGGVGVGFKPLANVGEASY---NY 164
Db 113 YALAGVGVKATFKESHSTQDDSFNSKISSRKTGFAMGAGVQMNPLENIIVDVGEGSNI 172
OY 165 LSEDANAISLGAHLAF 180
Db 173 SSTKINGFNVGVGRF 188

RESULT 11
US-08-475-749-5
Sequence 5, Application US/08475749
Patent No. 6010901
GENERAL INFORMATION:
APPLICANT: Miller III, Samuel I.
APPLICANT: Mekalanos, John L.
TITLE OF INVENTION: SALMONELLA VIRULENCE GENES
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P. C.
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 555X
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/475,749
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/090,526
FILING DATE: 09-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/629,602
FILING DATE: 18-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/192004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 188 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-475-749-5

Query Match 11.3%, Score 103; DB 3; Length 188;
Best Local Similarity 27.0%, Pred. No. 0.00071;
Matches 53; Conservative 28; Mismatches 91; Indels 24; Gaps 10;
OY 1 MKLTLLAVASASSLLASANAISYGSNADAPYVGAIKIGVDAKQINGKNTATAGIYAG 60
Db 1 MNITLTLVITTSVLVNVVAAQADTNAPSVGYARYAQSKV--QDFKNIRGVNVK----R 54
OY 61 YNFDO--NFGVEAEFVGSDAKEFNAGVSPV-----KGDVK--SFGAYGTYRYNINTPF 110
Db 55 YEDDSPVSPFSSLSLYLGD-RQASGSVEPEGIHYHDKFEVKKGSLMVGPAVRLS-DNFSL 112

OY 111 YAKKGLGIAT-KYVDVTSRNATTTYSNK--SDKTSLAGGVGVGFKPLANVGEASY---NY 164
Db 113 YALAGVGVKATFKESHSTQDDSFNSKISSRKTGFAMGAGVQMNPLENIIVDVGEGSNI 172
OY 165 LSEDANAISLGAHLAF 180
Db 173 SSTKINGFNVGVGRF 188

RESULT 12
US-08-200-232-2
Sequence 2, Application US/08200232
Patent No. 5721349
GENERAL INFORMATION:
APPLICANT: Cover, Timothy L.
APPLICANT: Blaser, Martin J.
TITLE OF INVENTION: VACUOLATING TOXIN-DEFICIENT H. PYLORI
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: NEEDLE & ROSENBERG P.C.
STREET: 127 Peachtree Street, Suite 1200
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/200,232
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Spratt, Gwendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 2200,023
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1287 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-200-232-2

Query Match 11.0%, Score 100; DB 1; Length 1287;
Best Local Similarity 24.8%, Pred. No. 0.022;
Matches 53; Conservative 32; Mismatches 71; Indels 58; Gaps 12;
OY 9 AVSASSLLAMGSANAISYGSNADAPYVGAIKIGVDA-----KQIN----- 49
Db 1025 AIGGAS-LNNGGNASL-YGTSGAVDAYLN---GQVEALIVGGSGSYGSSFNQANSLSNG 1079
OY 50 GKNTAYGIYAGINPDQN-FGYEAE-FVGSDAKEFNAGVSPYKGVKSF-----GA 97
Db 1080 ANNTNFGVYSRIFANQHFEDEADGALGSDSSLSLFXSALRLDLQSYNYLAYSAATRAS 1139
OY 98 YGYRYNFINTPFYAKKGLGIATKFTVDVTSRNATTTYSKSDPSTLAGGVGVGFKPLANVG 157
Db 1140 YG-IDFAFRNALVILKPSVGVSYNHLGSTNR-----SNSTKNVALSNSSSOHLFNASAN 1194
OY 158 VEASVNY-----LSE-----DANAISL 174
Db 1195 YEARYYYGDTSYFVNMAGVLDPEANFGSSNAVSL 1228

RESULT 13

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-178 <MIL>
A:Cross-references: GB:M29945; NTD:g155437; PIDN:AAA86894.1; PID:g155438
C:Superfamily: phage lambda membrane protein Iom

Query Match	13.5%	Score 123;	DB 2;	Length 178;
Best Local Similarity	29.6%;	Pred. No. 0.0018;		
Matches 55;	Conservative 29;	Mismatches 82;	Indels 20;	Gaps 9

[illegible]

```
RESULT      3
H81277
flagellin Cj1339c [imported] - Campylobacter jejuni (strain NCTC 11168)
```

C:Species: Campylobacter jejuni
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-May-2000
C:Accession: H81277
R:Packhill, J., Wren, B.W., Mungall, K., Kelley, J.M., Churcher, C., Basham, D., Chillif, C.W., Quail, M., Raddingham, M.A., Rutherford, K.M., VanVliet, A., Whitehead, S., Barrerford, A., 2000
N:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp
R:Reference number: H81250; MUID:20150912

Query Match	12.8%;	Score 117;	DB 2;	Length 572;
Best Local Similarity	24.5%;	Pred. No. 0.022;		
Matches	50;	Conservative	24;	Mismatches 70;
				Indels 60;
				Gaps 7

QY	22	AA	SVGSNADQ	PPV	VGKIGIGVADKQ-----	-INCKNTAYGIYA-----	59
		:	:	:	:	:	
Db	242	AAVRAGAT	STDFEAI	NGKIGV	DKDGDANGALVAALVAAVSKDTTGV	EASIDANGQLLTS	301
QY	60	----	GYNE	DONFGVEAE	FVSGDAKEFNAGVS	PKYGDVK-----	SFGATGYRY 103
Db	302	REGRCIK	IDGNI	GGCA-FINAD	MEKNTGRISLVK	NDCKDLILG	SNLSAGCFEA----- 354
QY	104	NEINT	PEYAKK	KLIGIA	IAKTR--VDVTS	RNAATYSSNKR	SDKTSLAG----- -GVGAFKPK 152
		:	:	:	:	:	:
Db	355	----	TOFISQ	ASVSLR	ESKQIDAN	INADMGFSG	ANKGVVLGGVSSVAYMSAGSGSFSS 410
QY	153	LANYG	VEASTY	TYLSE	DANALIS	LCA	176
		:	:	:	:	:	
Db	411	GS	GVSVSG	SKNYS	TGFAN	ALISA	434

RESULT
H82207

hypothetical protein VC1384 [imported] - *Vibrio cholerae* (group O1 strain N16961)
C:Species: *Vibrio cholerae*
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 15-Sep-2000
C:Accession: H82207
R:Heldebrery, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L., Dodson, R.,
Charidson, D., Ermolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Drysdale, I., Sellers,
I. R., Mekalanos, J.J., Venter, J.C., Fraser, C.M.

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-160 <HEI>
A:Cross-references: GB:AE004217; GB:AE003852; NID:96655866; PIDN:AAE94542.1; GSPDB:GNM
A:Experimental source: serogroup O1, strain N16561, biotype El Tor
C:Genetics:
A:Gene: VC1384
A:Map position: 1

Query Match	12.6%;	Score 114.5;	DB 2;	Length 160;
Best Local Similarity	29.8%;	Pred. No. 0.0084;		
Matches	56;	Conservative	19;	Mismatches 70;
				Indels 43;
				Gaps 9;

[illegible]

RESULT 5
D64050 major outer membrane protein P2 - Haemophilus influenzae (strain Rd KW20)
C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence-revision 18-Aug-1995 #text-change 21-Jul-2000
C:Accession: D64050
R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirschner, E.F.; Kerlavage
; Cocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman
; D.M.; Brannon, R.C.; Fine, L.D.; Fritchman, J.L.; Fullmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630
A:Accession: D64050
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-359 <TIGR>
A:Cross-references: GB:U02699; GB:U42023; NID:g93212180; PIDN:AAC21810.1; PID:g15730932
C:Keywords: membrane protein

```

Query Match      12.3%; Score 112.5; DB Z; Length 359;
Best Local Similarity 28.1%; Pred. No. 0.031;
Matches 52; Conservative 25; Mismatches 65; Indels 43; Gaps 11;

0Y      5 KTLIVASASLLAMASANAISYGSNADQPYVGARIGY-----DAKING-----K 51
      ||| | : | : |||| : : : : : : : : : : : : : : : : : : :
Db      3 KTLALILYAGFASANANAIVYNNEG--INVELGGRISIAEOSNTVDNQKQOHBALRNQ 61
      ||| | : | : |||| : : : : : : : : : : : : : : : : : :

0Y      52 NTAYGILYAGINDQNF-----GYAEAFV-----GSDAKDEFNAGSPVKGDVYSFGAYGTH 101
      ||| | : | : |||| : : : : : : : : : : : : : : : : : :

```


[illegible][illegible]

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-188 <PUI>
A:Cross-references: GB:M5546; NID:g154232; PIDN:AAA27179.1; PID:g154233
C:Superfamily: phage lambda membrane protein lom

Query Match 11.8%; Score 108; DB 2; Length 188;
Best Local Similarity 24.2%; Pred. No. 0.036;
Matches 50; Conservative 29; Mismatches 82; Indels 46; Gaps 8;

OY 1 MKTLTLLAVSSLLASANAISYNSADQPYVGAKEIGVDKOTNGKTAAGIYAG 60
DB 1 MKNITLSTLVITTSVLYVNAQADTNAFSVGAKYASQSKV--QDERKNIRGVNWKY----- 53
OY 61 YNFDONFGEAEFVGSDAKEFNAGVSPYKGDVKSFGA-----YGT----- 100
DB 54 -----RYEDDSPVSFISLSLYLXGDRQAGSVPEEGIHHDKFEVAYGSLMWGP 102
OY 101 -YRYFINTPFAKAGKLGIAK-TKVDYTSRNAITYSNK--SDKISLAGVGVEFKPLANV 156
DB 103 AYRLS-DNFSLYALAGVGVKATPEKHESTODSFSNKSISRKTGFAMCAGVQMNPLENI 161
OY 157 GVEASY---NYLSEADANAISGAHLAF 180
DB 162 VVDVGIEGSNISSTKINGFNVGVTFR 188

RESULT 9
MMBPP2
outer membrane porin lc precursor - phage PA2
C:Species: phage PA2
A:Note: host Escherichia coli
C:Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 02-Jul-1998
C:Accession: D25647
R:Blasband, A.J.; Marcotte Jr., W.R.; Schnaitman, C.A.
J. Biol. Chem. 261, 12723-12732, 1986
A:Title: Structure of the lc and mmpc outer membrane porin protein genes of lambdaoid bac
A:Reference number: A25647; MUID:86304457
A:Accession: D25647
A:Molecule type: DNA
A:Residues: 1-365 <BLA>
C:Genetics:
A:Gene: lc
C:Superfamily: outer membrane protein phoe
C:Keywords: membrane protein; porin; trimer
F:1-23/Domain: signal sequence #status predicted <SIG>
F:24-365/Product: outer membrane porin lc #status predicted <MAT>

Query Match 11.8%; Score 108; DB 1; Length 365;
Best Local Similarity 26.6%; Pred. No. 0.076;
Matches 62; Conservative 19; Mismatches 78; Indels 74; Gaps 11;

OY 1 MKTLKTL-AYSASSLLMSANAISY-----GNSADAQPYVGA 39
DB 1 MKKLVAISAVAASVLMASQAALTYNKSNDKLDLYGVNAKHFFSSNDADDGDTTYAR 60
OY 40 IGOVDAKINGKNFYAGYAGYFNPDONFGEAEFVGSDAKEFNAGVSPYKGDVKSFGAYG 99
DB 61 LGFKETQINDQLTGFQGW-YEFKGN---RAESGSSKDKTRLAFAGLK-----FGDYG 111
OY 100 TYRYNFINTPFAKGL-----GIATKTVDV-----TSRN----- 129
DB 112 SIDVGRNVAVDIGAWTDVLPFEGDPTWTQTDVEMTRGTGFAIYRRNDDPFLVDGLNF 171
OY 130 ATTYSNKSDKT---SLAGVGVEFKPLANVGVEASYNLSEADANAISGAHLA 179
DB 172 AAQYOGKNDKSDPDMYTEGNGDGF-----GFSATYEV-----EGFGIGATYA 213

RESULT 10
MMBCKC

outer membrane porin mmpc precursor - Escherichia coli cryptic lambdaoid prophage DUPI
C:Species: Escherichia coli
C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 16-Jul-1999
C:Accession: A25647; S66594; G64787
R:Blasband, A.J.; Marcotte Jr., W.R.; Schnaitman, C.A.

J. Biol. Chem. 261, 12723-12732, 1986
A:Title: Structure of the lc and mmpc outer membrane porin protein genes of lambdaoid
A:Reference number: A25647; MUID:86304457
A:Contents: mutant strain CS384
A:Accession: A25647

A:Molecule type: DNA
A:Residues: 1-365 <BLA>
R:Mandl, A.A.; Sharples, G.J.; Mandal, T.N.; Lloyd, R.G.
J. Mol. Biol. 257, 561-573, 1996

A:Title: Holiday junction resolvases encoded by homologous rusa genes in Escherichia
A:Reference number: S66579; MUID:96196428
A:Accession: S66594

A:Molecule type: DNA
A:Residues: 347-365 <MAH>

A:Cross-references: EMBL:X92587; NID:g1051136; PIDN:CAA63325.1; PID:g1051145
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
A.: Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617

A:Accession: G64787
A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA
A:Residues: 'NNIYRAVTSFENNSSKKGIT',1-325, 'N',327-347, 'RGAKKSIT', <BLAT>

A:Cross-references: GB:AE000160; GB:U000096; NID:g1786751; PIDN:AACT73654.1; PID:g17867
A:Experimental source: Strain K-12, substrain W61655
C:Comment: In wild-type strains of E. coli K-12, the mmpc open reading frame is inter
ain CS348, the IS5 element has been deleted and mmpc is expressed.

C:Genetics:
A:Gene: mmpc
A:Map position: 12 min
A:Genome: cryptic lambdaoid prophage DUPI2
C:Superfamily: outer membrane protein phoe
C:Keywords: membrane protein; porin; trimer
F:1-23/Domain: signal sequence #status predicted <SIG>
F:24-365/Product: outer membrane porin mmpc #status predicted <MAT>

Query Match 11.8%; Score 108; DB 1; Length 365;
Best Local Similarity 26.6%; Pred. No. 0.076;
Matches 62; Conservative 19; Mismatches 78; Indels 74; Gaps 11;

OY 1 MKTLKTL-AYSASSLLMSANAISY-----GNSADAQPYVGA 39
DB 1 MKKLVAISAVAASVLMASQAALTYNKSNDKLDLYGVNAKHFFSSNDADDGDTTYAR 60
OY 40 IGOVDAKINGKNFYAGYAGYFNPDONFGEAEFVGSDAKEFNAGVSPYKGDVKSFGAYG 99
DB 61 LGFKETQINDQLTGFQGW-YEFKGN---RAESGSSKDKTRLAFAGLK-----FGDYG 111
OY 100 TYRYNFINTPFAKGL-----GIATKTVDV-----TSRN----- 129
DB 112 SIDVGRNVAVDIGAWTDVLPFEGDPTWTQTDVEMTRGTGFAIYRRNDDPFLVDGLNF 171
OY 130 ATTYSNKSDKT---SLAGVGVEFKPLANVGVEASYNLSEADANAISGAHLA 179
DB 172 AAQYOGKNDKSDPDMYTEGNGDGF-----GFSATYEV-----EGFGIGATYA 213

RESULT 11
S68072
major outer membrane protein P2, antigenic variant t1 - Haemophilus influenzae
C:Species: Haemophilus influenzae
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 08-Oct-1999
C:Accession: S68072; S45437
R:Dum, B.
submitted to the EMBL Data Library, June 1993
A:Reference number: S68062

```
A:Accession: S68072
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-371 <DU1>
A:Cross-references: EMBL:X73386; NID:g510590; PIDN:CA51803.1; PID:g510591
A:Experimental source: isolate t1
R:Duim, B.; van Alphen, L.; Eljk, P.; Jansen, H.M.; Dankert, J.
Mol. Microbiol. 11, 1181-1189, 1994
A:Title: Antigenic drift of non-encapsulated Haemophilus influenzae major outer membrane
A:Reference number: S43699; MUID:94293786
A:Accession: S45437
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 174-188;215-234;257-283 <DUW>
A:Cross-references: EMBL:X73386
A:Experimental source: isolate t1
C:Keywords: membrane protein

Query Match          11.3%; Score 103; DB 2; Length 371;
Best Local Similarity 24.8%; Pred. No. 0.2;
Matches 53; Conservative 29; Mismatches 74; Indels 58; Gaps 12;

QY 5 KTLAASASSLLAMSAANAISYNSADAPYVAKIGQV-----DAKQING-----K 51
    |||.: : |:||||: | : : : : : : : : : : : : : : : : : : :
    3 KTLAALIVGAFASANAAMAVVYNNEG-SKVELGRLSVIAEOSNNTVDDQKQHGALRNQ 61
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 52 NTAYGIYAGYFDPQNF---GVEAEFVG---SPAKFENNGVSPVKGDVKSFGAYGYRYN 104
    : : | : | : | : | : | : | : | : | : | : | : | : | : | : |
    62 GSRFHAKATHNEDGFGYAGCYLFTFRIISHQDNADHFD-----DITF-----KYAVY 108
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 105 FINTPPYAKGLGIATKV-DVTSRNATYYS--NKSDKTSLAGG-VGVGFKPLANGVGA 160
    : : : : | : | : | : | : | : | : | : | : | : | : | : | :
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 109 TLGNKAFGEVKLGRAKTIADDTISADKEGYGLNNSKYIRTNNGTYGYFFKQID--GLVL 166
QY 161 SYNYLSEDA-----MAISLGA 176
    |||.: : |:||||: | : : : : : : : : : : : : : : : : : : :
DB 167 GANYLLAQAARDTANPCKGEVAQAOSISNGVOGA 200

RESULT 12
S68069
Major outer membrane protein P2, antigenic variant t2 - Haemophilus influenzae
C:Species: Haemophilus influenzae
C:Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 08-Oct-1999
A:Accession: S68069; S43706
R:Duim, B.
submitted to the EMBL Data Library, June 1993
A:Reference number: S68062
A:Accession: S68069
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-371 <DU1>
A:Cross-references: EMBL:X73383; NID:g475192; PIDN:CA51800.1; PID:e1192128; PID:g265429
A:Experimental source: isolate t2
R:Duim, B.; van Alphen, L.; Eljk, P.; Jansen, H.M.; Dankert, J.
Mol. Microbiol. 11, 1181-1189, 1994
A:Title: Antigenic drift of non-encapsulated Haemophilus influenzae major outer membrane
A:Reference number: S43699; MUID:94293786
A:Accession: S43706
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 174-188;215-234;257-283 <DUW>
A:Cross-references: EMBL:X73383
A:Experimental source: isolate t2
C:Keywords: membrane protein

Query Match          11.3%; Score 103; DB 2; Length 371;
Best Local Similarity 24.8%; Pred. No. 0.2;
Matches 53; Conservative 29; Mismatches 74; Indels 58; Gaps 12;
```

```
DB 3 KTLAALIVGAFASANAAMAVVYNNEG-SKVELGRLSVIAEOSNNTVDDQKQHGALRNQ 61
QY 52 NTAYGIYAGYFDPQNF---GVEAEFVG---SPAKFENNGVSPVKGDVKSFGAYGYRYN 104
    : : | : | : | : | : | : | : | : | : | : | : | : | : | : |
    62 GSRFHAKATHNEDGFGYAGCYLFTFRIISHQDNADHFD-----DITF-----KYAVY 108
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 105 FINTPPYAKGLGIATKV-DVTSRNATYYS--NKSDKTSLAGG-VGVGFKPLANGVGA 160
    : : : : | : | : | : | : | : | : | : | : | : | : | : | :
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 109 TLGNKAFGEVKLGRAKTIADDTISADKEGYGLNNSKYIRTNNGTYGYFFKQID--GLVL 166
QY 161 SYNYLSEDA-----MAISLGA 176
    |||.: : |:||||: | : : : : : : : : : : : : : : : : : : :
DB 167 GANYLLAQAARDTANPCKGEVAQAOSISNGVOGA 200

RESULT 13
A56152
Major 25k outer membrane protein precursor - Brucella abortus
C:Species: Brucella abortus
C:Date: 03-Oct-1995 #sequence_revision 03-Oct-1995 #text_change 08-Oct-1999
A:Accession: A56152
R:de Wergifosse, P.; Lintermans, P.; Limelet, J.N.; Cloeckaert, A.
J. Bacteriol. 177, 1911-1914, 1995
A:Title: Cloning and nucleotide sequence of the gene coding for the major 25-kilodalton
A:Reference number: A56152; MUID:95204367
A:Accession: A56152
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-213 <DEA>
A:Cross-references: GB:X79284; NID:g769744; PIDN:CA55872.1; PID:g769745

Query Match          11.2%; Score 102.5; DB 2; Length 213;
Best Local Similarity 27.1%; Pred. No. 0.12;
Matches 60; Conservative 22; Mismatches 60; Indels 79; Gaps 11;

QY 1 MKTKTLAVSASSLLAMSAANAISYNSADA---OPYGAKI----- 40
    |:|:|:|: | : | : | : | : | : | : | : | : | : | : | : | :
DB 1 MRTLKSLVIVSA-ALLPFSATAF-----AADAIQEQQPPVPAPVAVAPQYSMAAGTYGLYL 54
QY 41 --GQVDAK-----QINGKNTAYGIYAGYFDPN---FGVEAEFVGSDAKFENAGVSPVK 90
    : : : : | : | : | : | : | : | : | : | : | : | : | : | :
    55 GYGMNKAKTSTVGSIKPDDMKAGAFAGWNPFOQIYGVGDAGYSMAKSKDGLVKG 114
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 91 DVKSPFAGCYRYRNFINTPPYAKGKIGT-----AKTKVY 125
    : : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 115 FEGSLGARVGYDLNPV-MPYLTAAGIASQIKLNNGLDDESKFRVGTAGAGLEAKLTLDNI 173
QY 126 TSR---NATYYSNKSDDKTSLAG-----GVGVGFK 151
    | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 174 LGRVEYRYTYQGNKN--YDLAGTVYNNKLDTDQDIRVGIYK 212

RESULT 14
S68062
Major outer membrane protein P2 - Haemophilus influenzae (isolate d2 and isolate 3dR)
C:Species: Haemophilus influenzae
A:Variety: isolate d2; isolate 3dR
C:Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 08-Oct-1999
A:Accession: S68062; S68065; S43699
R:Duim, B.
submitted to the EMBL Data Library, June 1993
A:Reference number: S68062
A:Accession: S68062
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-386 <DU1>
A:Cross-references: EMBL:X73376; NID:g475178; PIDN:CA51793.1; PID:e1192121; PID:g265
A:Experimental source: isolate d2
A:Accession: S68065
A:Status: preliminary
A:Molecule type: DNA
```



```
Db 51 VGYAFDKNRLAVDTYNGKVTANADVVDVSLKSGSLGTGFDPDLADFKPY--GVR 108
OY 118 IAKTVVDYTSRNATYYSKSKDTSIAGV---GVGEKPLANVGEAS--YNYISSEDANA 171
Db 109 VSTNCADYTA-NARYRIEAFATEFRIGIGALAGVQYKLTQVVAINTNIEYRLASNVSD 167
OY 172 ISLGAHLAF 180
Db 168 VGKAGLRP 176

RESULT 2
ID AIL_YEREN STANDARD: PRT: 178 AA.
AC P16454;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ATTACHMENT INVASION LOCUS PROTEIN PRECURSOR.
GN AIL.
OS Yersinia enterocolitica.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
CC Yersinia.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 24-33.
RX MEDLINE-90130261; PubMed-1688838;
RA Miller V.L., Bliska J.B., Falkow S.;
RT "Nucleotide sequence of the Yersinia enterocolitica ail gene and
RT characterization of the Ail protein product."
RL J. Bacteriol. 172:1062-1069(1990).
CC -1- FUNCTION: THIS MEMBRANE-ASSOCIATED PROTEIN PROMOTES INVASION
CC OF PATHOGENIC BACTERIA INTO EUKARYOTIC CELLS BY AN UNKNOWN
CC MECHANISM.
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE AIL/OMPX/PACG/LDM FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, M29945; AAA8694.1; -.
CC PIR, A5123; A5123.
CC INTERPRO: IPR000758; -.
CC PRINTS: PR00316; ENTEROVIROMP.
CC PROSITE: PS00694; ENT_VIR_OMP_1; 1.
CC PROSITE: PS00695; ENT_VIR_OMP_2; 1.
CC Outer membrane; Transmembrane; Signal; Virulence.
CC SIGNAL 1 23
CC CHAIN 24 178 ATTACHMENT INVASION LOCUS PROTEIN.
CC FT 178 AA; 19548 MW; D46FC92E2524F162 CRC64;
CC SEQUENCE

Query Match 13.5%, Score 123; DB 1; Length 178;
Best Local Similarity 29.6%; Pred. No. 0.00091;
Matches 55; Conservative 29; Mismatches 82; Indels 20; Gaps 9;
```

```
Db 173 GAGYRF 178

RESULT 3
ID FLAI_CAMJE STANDARD: PRT: 571 AA.
AC P56963; Q9PMW0;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE FLAGELLIN A.
GN FLA OR C11339C.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
CC Campylobacter.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NCTC 11168;
RX MEDLINE-20150912; PubMed-10688204;
RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feltham T., Holtroyd S.,
RA Jagers K., Kariyasev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Raftery M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrett B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences."
RL Nature 403:665-668(2000).
CC -1- FUNCTION: FLAGELLIN IS THE SUBUNIT PROTEIN WHICH POLYMERIZES TO
CC FORM THE FILAMENTS OF BACTERIAL FLAGELLA.
CC -1- SUBUNIT: HETEROPOLYMER OF FLA AND FLAB.
CC -1- SIMILARITY: BELONGS TO THE BACTERIAL FLAGELLIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, AL139078; CAB3766.1; -.
CC KW Flagella.
CC FT INT.MET 0 0 BY SIMILARITY.
CC FT SEQUENCE 571 AA; 58907 MW; 9C63B2A10C1AE863 CRC64;
CC SEQUENCE

Query Match 12.8%, Score 117; DB 1; Length 571;
Best Local Similarity 24.5%; Pred. No. 0.01;
Matches 50; Conservative 24; Mismatches 70; Indels 60; Gaps 7;
```

```

DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ATTACHMENT INVASION LOCUS PROTEIN PRECURSOR.
GN ALL.
OS Yersinia pseudotuberculosis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-YPIII:
RX MEDLINE-96294755; PubMed-86698470;
RA Yang Y., Herriam J.J., Mueller J.P., Isberg R.R.;
RA "The psa locus is responsible for the noninducible binding of Yersinia
RT pseudotuberculosis to cultured cells."
RL Infect. Immun. 64:2483-2489(1996).
CC -1- FUNCTION: THIS MEMBRANE-ASSOCIATED PROTEIN PROMOTES INVASION
CC OF PATHOGENIC BACTERIA INTO EUKARYOTIC CELLS BY AN UNKNOWN
CC MECHANISM (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE AIL/OMPX/PAGC/LOW FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: L9439; AAB36601.1; -
CC INTERPRO: IPR000758; -
CC PRINTS: PRO0316; ENTEROVIROMP.
CC PROSITE: PS00694; ENT_VIR_OMP_1: 1.
CC PROSITE: PS00695; ENT_VIR_OMP_2: 1.
CC Outer membrane; Transmembrane; Signal; Virulence.
CC SIGNAL
FT CHAIN 1 26 BY SIMILARITY.
FT SEQUENCE 182 AA; 20192 MW; EET7A4A20E4D8975C CRC64;
SQ
Query Match 12.7%; Score 115.5; DB 1; Length 182;
Best Local Similarity 27.8%; Pred. No. 0.004;
Matches 52; Conservative 29; Mismatches 85; Indels 21; Gaps 7;
QY 5 KTLAVSASSLLAMSAANAISYGSNADAPYGAIGQVADAKOINKNATYGIACYND 64
DB 6 KTLVSSLACUSIASVNVYABESSISIGYASRV-KDDGYKLDNPNPGFNKRYEEN 64
QY 65 QNEGVAEFEVGSDA-----KEFNAGVSPYKGDYK--SFGAYGYRYNFIINTPEYAKGL 116
DB 65 NDMGV-----IGSFAQRRRGESVDGFKLIDGPKYYSYAGVFR--INEVSLXGL 117
QY 117 GAKTVVDVTSNATYYSNKSQDTSIAGVGNGVFKPLANGVNASYNSSEDANAIS--- 173
DB 118 GAGHGKRAKVS--IFGQSESRKSTSLAGLOFNPHEVIDASYEYSKLDVKGKTWM 175
QY 174 LGAHLP 180
DB 176 LGAGYRF 182

```

```

RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN-WTHI 1128;
RX MEDLINE-9422575; PubMed-7909539;
RA Sirakova T., Kolatukudy P.E., Murvin D., Billy J., Leake E.,
RA Lim D., Demaria T., Bakaletz L.;
RA "Role of fimbriae expressed by nontypeable Haemophilus influenzae in
RT pathogenesis of and protection against otitis media and relatedness
RT of the fimbria subunit to outer membrane protein A."
RL Infect. Immun. 62:2002-2020(1994)
CC -1- FUNCTION: ACTS AS A FIMBRIAE SUBUNIT.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE OMPA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: L08448; AAA24959.1; -
CC INTERPRO: IPR000498; -
CC PRINTS: PRO0145; -
CC PFAM: PF00691; OMPA: 1.
CC PFAM: PF01389; OMPA.membrane: 1.
CC PRINTS: PRO1021; OMPADOMAIN.
CC PROSITE: PS01068; OMPA; FALSE_NEG.
CC Outer membrane; Transmembrane; Porin; Signal; Fimbria.
CC SIGNAL
FT CHAIN 1 21
FT DISULFID 332 359 OUTER MEMBRANE PROTEIN P5.
FT DOMAIN 332 344 BY SIMILARITY.
FT SEQUENCE 359 AA; 38340 MW; 576B1C59B4818C37 CRC64;
SQ
Query Match 12.4%; Score 113; DB 1; Length 359;
Best Local Similarity 28.2%; Pred. No. 0.013;
Matches 55; Conservative 18; Mismatches 80; Indels 42; Gaps 11;
QY 5 KTLAVSASSLLAMSAANAISYGSNADAPYGAIGQ--VDAKOING----- 50
DB 3 KTAIVLVAGLAASVAQAPOENTF---YAGVKAGGSGFHGINNAIKKGLSSNV 58
QY 51 ---KNT-ANGTAYGYNE--DQNGVAEFEVGSQ---AKEFNAGVSPYKGDYKSGFAY--- 98
DB 59 GYRNMTFTYGVFEGYQIILNDNFGLAELGYDDFGRAKLEAGKPRAKH--TNHGAVLSL 116
QY 99 -GTYRYNFIPTPYAKGKGIATKYD--VTSRNATYYSNKSQDTSIAGG---VGVGFK 151
DB 117 KGSYE---VLDGLDVGKACVALVRSYKFEEDANGTRHKKRHHYARASGLPAVGAETA 173
QY 152 PLANVGEASYNLYS 166
DB 174 VLPELAVRLEYQWLT 188

```

```

RESULT 5
OM53_HAEIN STANDARD; PRT; 359 AA.
AC P45936;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE OUTER MEMBRANE PROTEIN P5 PRECURSOR (OMP P5) (FIMBRIIN).
GN OMPA OR OMP5.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.

```

```

RESULT 6
OM21_HAEIN STANDARD; PRT; 359 AA.
AC P43839;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE OUTER MEMBRANE PROTEIN P2 PRECURSOR (OMP P2).
GN OMP2 OR H10139.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RD / KW20;

```

RX MEDLINE-95350630; PubMed-7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McElveney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shiley R., Liu L.-I., Glodet A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uitterback T.R., Hanna M.C., Nguyen D.T., Saudak D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghegan N.S.M.,
RA Gench C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd. ";
RL Science 269:496-512(1995).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
CC -1- SIMILARITY: TO PHOTOBACTERIUM STRAIN SS9 OMPH.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U32699; AAC21010.1; -.
DR TIGR: H10139; -.
KW Outer membrane; Transmembrane; Porin; Signal.
KW SIGNAL 1 20 BY SIMILARITY.
FT CHAIN 21 359 OUTER MEMBRANE PROTEIN P2.
SQ SEQUENCE 359 AA; 39375 MW; 7E9E9FEC2BD50314 CRC64;

Query Match 12.3%; Score 112.5; DB 1; Length 359;
Best Local Similarity 28.1%; Pred. No. 0.015;
Matches 52; Conservative 25; Mismatches 65; Indels 43; Gaps 11;

QY 5 KTLAVSASSLLAMANAISYGSNADAPYGAIGOV-----DAKQING-----K 51
DB 3 KTLAALIYGAFAPASANAAMAVYNNEG-TNVELGRLSTIAEQSNSTVDNOKOQGLARNQ 61
QY 52 NTAYGIYAGVYNDQNF---GVEAEFV-----GSDAKEFNAGVSPVKGDFKSPGAGCTY 101
DB 62 GSRFHAKTHNFGDGFYAGYLETREFTVKASNGSD--NF-----GDITS-----KY 106
QY 102 RYNEINTPEYAKGLGIATKYVD-VTSRNATYTS--NKSDKTSLAGG-VGVGFKPLANVG 157
DB 107 AYVTLGNKNAFGEVKLGRAKTIADGITSADKEYGYLVNNSDYIPTSGNTVGYTFKIDGLV 166
QY 158 YEAST 162
DB 167 LGANY 171

RESULT 7
OM22_HAEIN STANDARD; PRT; 361 AA.
ID OM22_HAEIN
AC P20149;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE OUTER MEMBRANE PROTEIN P2 PRECURSOR (OMP P2).
GN OMP2.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SEROTYPE B;
RX MEDLINE-89079316; PubMed-2535836;
RT Munson R.S. Jr., Toian R.W. Jr.;
RT "Molecular cloning, expression, and primary sequence of outer
RT membrane protein P2 of Haemophilus influenzae type b. ";
RL Infect. Immun. 57:88-94(1989).

RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-SEROTYPE B;
RX MEDLINE-90158127; PubMed-2576096;
RA Munson R.S. Jr., Bailey C., Grass S.;
RT "Diversity of the outer membrane protein P2 gene from major clones of
RT Haemophilus influenzae type b. ";
RL Mol. Microbiol. 3:1797-1803(1989).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-SEROTYPE B;
RX MEDLINE-89173305; PubMed-2538396;
RA Hansen E.J., Hasemann C., Clausell A., Capra J.D., Orth K.,
RA Woonaw C.R., Slaughter C.A., Latimer J.D., Miller E.E.;
RT "Primary structure of the porin protein of Haemophilus influenzae
RT type b determined by nucleotide sequence analysis. ";
RL Infect. Immun. 57:1100-1107(1989).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
CC -1- SIMILARITY: TO PHOTOBACTERIUM STRAIN SS9 OMPH.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: J03359; AAA24993.1; -.
DR EMBL: A09003; CA00819.1; -.
DR PIR: A30542; A30542.
DR PIR: S09622; S09622.
KW Outer membrane; Transmembrane; Porin; Signal.
KW SIGNAL 1 20
FT CHAIN 21 361 OUTER MEMBRANE PROTEIN P2.
SQ SEQUENCE 361 AA; 39701 MW; AF104C9CC08942D69 CRC64;

Query Match 12.3%; Score 112.5; DB 1; Length 361;
Best Local Similarity 28.1%; Pred. No. 0.015;
Matches 52; Conservative 25; Mismatches 65; Indels 43; Gaps 11;

QY 5 KTLAVSASSLLAMANAISYGSNADAPYGAIGOV-----DAKQING-----K 51
DB 3 KTLAALIYGAFAPASANAAMAVYNNEG-TNVELGRLSTIAEQSNSTVDNOKOQGLARNQ 61
QY 52 NTAYGIYAGVYNDQNF---GVEAEFV-----GSDAKEFNAGVSPVKGDFKSPGAGCTY 101
DB 62 GSRFHAKTHNFGDGFYAGYLETREFTVKASNGSD--NF-----GDITS-----KY 106
QY 102 RYNEINTPEYAKGLGIATKYVD-VTSRNATYTS--NKSDKTSLAGG-VGVGFKPLANVG 157
DB 107 AYVTLGNKNAFGEVKLGRAKTIADGITSADKEYGYLVNNSDYIPTSGNTVGYTFKIDGLV 166
QY 158 YEAST 162
DB 167 LGANY 171

RESULT 8
PORT_BPPA2 STANDARD; PRT; 365 AA.
ID PORT_BPPA2
AC P07238;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE OUTER MEMBRANE PORIN PROTEIN LC PRECURSOR.
GN LC.
OS Bacteriophage PA-2.
OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-86304457; PubMed-3017988;

RA Blasband A.J., Marcotte W.R. Jr., Schnaitman C.A.;
 RT "Structure of the 1c and nmcp outer membrane porin protein genes of
 RT lambdaoid bacteriophage."
 RT J. Biol. Chem. 261:12723-12732(1986).
 CC -1- FUNCTION: PORINS ARE MAJOR PROTEINS FOUND IN THE OUTER MEMBRANES
 CC OF GRAM-NEGATIVE BACTERIA WHERE THEY FORM CHANNELS FOR THE
 CC NONSPECIFIC PERMEATION OF SMALL SOLUTES (MOLECULES WITH MW LOWER
 CC THAN 4000-6000 DALTONS).
 CC -1- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE
 CC (BY SIMILARITY).
 CC -1- MISCELLANEOUS: A PORIN GENE CAN ALSO BE FOUND IN THE GENOMES OF
 CC CERTAIN LAMBDOID BACTERIOPHAGE, AND ITS PROTEIN IS EXPRESSED IN
 CC THE LYSOGENIC STATE. IN E. COLI THE EXPRESSION OF OMPC AND OMPF
 CC PROTEINS IS THEN REDUCED SUBSTANTIALLY.
 CC -1- SIMILARITY: BELONGS TO THE OMPC/PHO FAMILY OF PORINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: J02580; AAA32301.1; -.
 DR PIR: D25647; MMBPP2.
 DR HSSP: P02931; 1GFN.
 DR INTERPRO: IPR001702; -.
 DR INTERPRO: IPR001897; -.
 DR PFAM: PF00267; Gram-ve_porins; 1.
 DR PRINTS: PR00182; ECOLI_PORIN.
 DR PRINTS: PR00183; ECOLI_PORIN.
 DR PROSITE: PS00576; GRAM_NEG_PORIN; 1.
 KM Outer membrane; Transmembrane; Porin; Signal.
 FT SIGNAL 1 23
 FT CHAIN 24 365 OUTER MEMBRANE PORIN PROTEIN LC.
 FT CONFLICT 99 99 H -> R (IN AA SEQUENCE).
 FT SEQUENCE 365 AA; 40290 MW; 0FBC0531FB9C0205 CRC64;
 SQ
 Query Match 12.0%; Score 109; DB 1; Length 365;
 Best Local Similarity 26.6%; Pred. No. 0.029;
 Matches 62; Conservative 20; Mismatches 77; Indels 74; Gaps 11;
 QY 1 MKTLKTL-AYGASSLAMSANAISY-----GNSADAQPIYGA 39
 DB 1 MKRLVAISAVAASVLAAMSQAALTYNKRDSNKLDTGKYNKHYSSNDADGDTTYAR 60
 QY 40 IGVDAKQJNGKNTAGIAGYNFDQNFGEAEFVSGDAKEFNAGVSPKGDVKSFGAG 99
 DB 61 LQFKGTQINDQLTGEGQVEYFKGN---RAESQSSSKDKTHLAFAGLK-----FGDYG 111
 QY 100 TYRYNFIPTFYAKGL-----GIATKYDV-----TSRN----- 129
 DB 112 SIDYGRNYGVAVDYGAMTVDLPEFGDWTQIDVFWTGRGTGATYRNDFGLVDGLNF 171
 QY 130 ATTSYKSKSKT---SLAGVGVGFKPLANVGEASINYISEDAANAISLGAHLA 179
 DB 172 AAQYQKNDKSDPDNYTEGNGDGF-----GFSATYEV-----EGFGIGATYA 213
 RESULT 9
 PAGC_SALTY STANDARD: PRT; 188 AA.
 AC P23988;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE VIRULENCE MEMBRANE PROTEIN PAGC PRECURSOR.
 GN PAGC.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Salmonella.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 14028;
 RX MEDLINE=91100323; PubMed=1846140;
 RA Pulkkinen W.S., Miller S.L.;
 RT "A Salmonella typhimurium virulence protein is similar to a Yersinia
 RT enterocolitica invasion protein and a bacteriophage lambda outer
 RT membrane protein."
 RT J. Bacteriol. 173:86-93(1991).
 CC -1- FUNCTION: ESSENTIAL FOR FULL VIRULENCE AND SURVIVAL WITHIN
 CC MACROPHAGES.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
 CC -1- SIMILARITY: BELONGS TO THE AIL/OMPX/PAGC/LOW FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M55546; AAA27179.1; -.
 DR PIR: A39185; A39185.
 DR STGENE: SG10676; PAGC.
 DR INTERPRO: IPR000758; -.
 DR PRINTS: PR00316; ENTEROVIROMP.
 DR PROSITE: PS00694; ENT_VIR_OMP_1; 1.
 DR PROSITE: PS00695; ENT_VIR_OMP_2; 1.
 KM Outer membrane; Transmembrane; Signal; Virulence.
 FT SIGNAL 1 23
 FT CHAIN 24 188 VIRULENCE MEMBRANE PROTEIN PAGC.
 FT SEQUENCE 188 AA; 20574 MW; B23826F42B62DDFE CRC64;
 SQ
 Query Match 11.8%; Score 108; DB 1; Length 188;
 Best Local Similarity 24.2%; Pred. No. 0.017;
 Matches 50; Conservative 29; Mismatches 82; Indels 46; Gaps 8;
 QY 1 MKTLKTLAVSASSLLAMSANAISYGSNSADAQPIYGANIGOVDAKQINGKNTAYIAG 60
 DB 1 MKNIIISTVITTSYLVVNAQADTNAFSGVARYAQSKV--ODFKNIGVNVKY----- 53
 QY 61 YNDQNFGEAEFVSGDAKEFNAGVSPVKGDVKSFEA-----YGT----- 100
 DB 54 -----RYEDDSPVSFISLSLYLGRQASGSVEPEGIHYHDKFEYKYSGLMGP 102
 QY 101 -YRYNFIPTFYAKGLGIAGL-TKVDVTSRNATYTSNK--SDKTSIAGVGVGFKPLANV 156
 DB 103 AYRLS-DNFSIYLAAGVGIVKATFKRHSYQDGDSPSNKISSKRTGTGAWGAGVOMPLENI 161
 QY 157 GVEASY---NYLSEDAANAISLGAHLAF 180
 DB 162 VVDVGYEGSNISSITKINGFNVGGYRF 188
 RESULT 10
 NMPC_ECOLI STANDARD: PRT; 365 AA.
 AC P21420; P77189;
 DT 01-MAY-1991 (Rel. 18, Created)
 DT 01-MAY-1991 (Rel. 18, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE OUTER MEMBRANE PORIN PROTEIN NMPC PRECURSOR.
 GN NMPC OR PHMA.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MUTANT CS384;
 RX MEDLINE=86304457; PubMed=3017988;

RA Blasband A.J., Marcotte W.R. Jr., Schnaltman C.A.;
 RT "Structure of the Jc and mmpc outer membrane porin protein genes of
 RL lambdaoid bacteriophage.";
 RN J. Biol. Chem. 261:12723-12732(1986).
 [12]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE-97426617; PubMed-9278503;
 RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 [13]
 RP SEQUENCE FROM N.A.
 RA Chung E., Allen E., Araujo R., Aparicio A., Davis K., Duncan M.,
 RA Federapfel N., Hyman R., Kalman S., Komp C., Kurdi O., Lew H., Lin D.,
 RA Naneish A., Oefner P., Roberts D., Schramm S., Davis R.W.;
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 [14]
 RP SEQUENCE OF 1-17 FROM N.A.
 RC STRAIN-JL5502;
 RX MEDLINE-94335635; PubMed-8057841;
 RA Coll J.L., Heyde M., Portetier R.;
 RT "Expression of the mmpc gene of Escherichia coli K-12 is modulated by
 RT external pH. Identification of cis-acting regulatory sequences
 RL Mol. Microbiol. 12:83-93(1994).
 [15]
 RP SEQUENCE OF 347-365 FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE-96196428; PubMed-8648624;
 RA Mandi A.A., Sharples G.J., Mandal T.N., Lloyd R.G.;
 RT "Holiday junction resolvases encoded by homologous rusa genes in
 RT Escherichia coli K-12 and phage 82.";
 RL J. Mol. Biol. 257:561-573(1996).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
 CC -1- MISCELLANEOUS: IN WILD-TYPE STRAINS OF E. COLI K12, THE NMPC OPEN
 CC READING FRAME IS INTERRUPTED BY AN IS5 INSERTION AND GENERATES A
 CC HYBRID OPEN READING FRAME THAT IS NOT EXPRESSED. HOWEVER, IN
 CC MUTANT STRAIN CS348, THE IS5 ELEMENT HAS BEEN DELETED AND NMPC IS
 CC EXPRESSED.
 CC -1- SIMILARITY: BELONGS TO THE OMPC/PHO FAMILY OF PORINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M13457; AAA23728.1; ALT_SEQ.
 DR EMBL: AE000160; AAC73654.1; ALT_SEQ.
 DR EMBL: U82598; AAB40749.1; ALT_SEQ.
 DR EMBL: Z35442; CA884594.1; -.
 DR EMBL: X92587; CA63325.1; -.
 DR PIR: A25647; MMECNC.
 DR HSSP: P02931; IGFN.
 DR ECOGENE: EG10659; NMPC.
 DR INTERPRO: IPR001702; -.
 DR INTERPRO: IPR001897; -.
 DR PFM: PF00267; Gram-ve--porins; 1.
 DR PRINTS: PRO0182; ECOLNEIPORIN.
 DR PRINTS: PRO0183; ECOLNEIPORIN.
 DR PROSITE: PS00576; GRAM_NEG_PORIN; 1.
 KW Outer membrane; transmembrane; Porin; Signal.
 FT SIGNAL 1 23
 FT CHAIN 24 365 OUTER MEMBRANE PORIN PROTEIN NMPC.
 FT CONFLICT 326 326 K -> N (IN REF. 2 AND 3).
 FT SEQUENCE 365 AA; 40316 MW; 6E5128D4847FB4F8 CRC64;

Query Match 11.8%; Score 108; DB 1; Length 365;
 Best Local Similarity 26.6%; Pred. No. 0.036;
 Matches 62; Conservative 19; Mismatches 78; Indels 74; Gaps 11;
 QY 1 MKTLKTL-AYSASSLLMSANAISY-----GNSDAQPYVCAK 39
 DB 1 MKTLVAISAVAAVSILMAASAIAIYKNSKLDLYGVNAKHFFSSNDADGDTYAR 60
 QY 40 IGQVAKOINGKNTFYAGYNFDQNGFVEAFVGSAPKEFNAGVSPVKDDVSFGAYG 99
 DB 61 LGFKETQINQLQIFGQWE-YEFKGN---RAESGSSKDKTRLAFAGLK-----FGDYG 111
 QY 100 TYRYNFINTPYAKGL-----GIAKTYDV-----TSRN----- 129
 DB 112 SIDYGRNAGVAYDIGAMTDVLPFEFGDTWTQDTVPMTORATGVATRYRNNDFGLVDLNF 171
 QY 130 APTYSNKSQDKT---SLAGVGVEFPLNAGVEASYNLSLSDANAISGAHLA 179
 DB 172 AAQYQKNDKSDPDMYTEGNQGF-----GFSATYEX-----EGFGIGATYA 213
 RESULT 11
 OM31_BRUME STANDARD; PRT; 240 AA.
 ID OM31_BRUME
 AC Q45322;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE 31 KDA OUTER-MEMBRANE IMMUNOGENIC PROTEIN PRECURSOR.
 GN OMP31.
 OS Brucella melitensis.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Brucellaceae; Brucella.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-16M;
 RX MEDLINE-96355886; PubMed-8751924;
 RA Viccaino N., Cloeckaert A., Zygmunt M.S., Dubray G.;
 RT "Cloning, nucleotide sequence, and expression of the Brucella
 RT melitensis omp31 gene coding for an immunogenic major outer membrane
 RT protein.";
 RL Infect. Immun. 64:3744-3751(1996).
 CC -1- FUNCTION: MAJOR OUTER MEMBRANE PROTEIN ASSOCIATED WITH
 CC PEPTIDOGLYCAN. MAY FUNCTION AS A PORIN.
 CC -1- SUBUNIT: OLIGOMERIC.
 CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE.
 CC -1- SIMILARITY: BELONGS TO THE OMP25 / OMPB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF076290; AAB36693.1; -.
 DR INTERPRO: IPR000498; -.
 DR PFM: PF01389; Ompa_membrane; 1.
 KW Antigen; Outer membrane; Porin; Signal.
 FT SIGNAL 1 19
 FT CHAIN 20 240 POTENTIAL.
 FT DOMAIN 48 83 31 KDA OUTER-MEMBRANE IMMUNOGENIC
 FT PROTEIN.
 FT EPTROPE RECOGNIZED BY THE MONOCLONAL
 FT ANTI-BODY A59/10F09/G10.
 FT SEQUENCE 240 AA; 25323 MW; 21C65EC479F66A5 CRC64;

Query Match 11.5%; Score 105; DB 1; Length 240;
 Best Local Similarity 24.4%; Pred. No. 0.041;
 Matches 51; Conservative 29; Mismatches 85; Indels 44; Gaps 9;

```
OY 4 KTLTLLAVSASSLLAMSAANA---ISYGNADAP-----YVGAKIG----- 41
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 MASYLLASIAMFATSAMAADYVVSSEPSAPYDFTSWGTGYIGIYAGGKFKHPF 60
OY 42 ----GVDAKOLNGK-----NTAVGIYAGYNDQNGV-----EAEFVGSDAK-ERNAGVS 86
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 61 SSPDEDNDQVSGSLDVYAGGFGVGVQAGYMWQDNGVGLAETDFQSSVTGSIASAGS 120
OY 87 PVKGDVKS-FGAYGTYRYNFINTP-----FYAKGKLGIAKTVDVT-SKNATTYSNKSNDK 139
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 121 GLEGAEKFEKVEFGTVRRRLGTATERLMTVYGTGGLAYGKASAINLGDGASALHTMSDK 180
OY 140 TSLAGVGVGFKPLANVGVEASYNVLSLD 168
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 181 TKAGWTLGAGAEYAINNMWTLKEXLYTD 209
```

```
RESULT 12
OM2B_HAEIN STANDARD: PRT; 363 AA.
AC Q48220:
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE OUTER MEMBRANE PROTEIN P2 PRECURSOR (OMP P2).
GN OMP2.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
CC Haemophilus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3230B;
RX MEDLINE=94018553; PubMed=8412618;
RA Duijn B., Dankert J., Jansen H.M., van Alphen L.;
RT "Genetic analysis of the diversity in outer membrane protein P2 of
RT non-encapsulated Haemophilus influenzae.";
RL Microb. Pathog. 14:451-462(1993).
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC CC
DR EMBL; X73391; CA51808.1; -.
KW Outer membrane; Transmembrane; Porin; Signal.
FT SIGNAL 1 20
FT CHAIN 21 363 OUTER MEMBRANE PROTEIN P2.
FT FT
SQ SEQUENCE 363 AA; 39873 MW; 2739FD0BF42A3051 CRC64;
```

```
Query Match 11.5%; Score 105; DB 1; Length 363;
Best Local Similarity 26.9%; Pred. No. 0.063;
Matches 52; Conservative 29; Mismatches 78; Indels 34; Gaps 11;
```

```
OY 5 KTLTLLAVSASSLLAMSAANAISYGNADAPYVGAKIGY-----DAKQING-----K 51
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3 KTLAALITVGAFAASANAANAAYVYNNEG-TNVELGGRISLIAEOSNSTIKKQKQHGALRNQ 61
OY 52 NTAVGIYAGYNDQNF-----GVEAEFVGSDAKEFNAGVSPVKGVSFGAYGT-YRYNFI 106
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 62 SSRFHFKATHNFGDGYAGYLETRELVASQ-----SGTESD--NFGHITTKAYVYL 111
OY 107 NTPPYAKGKLGIAKTKVD-VTSRNATTYS--NKSQDTSLAGG-VGVGFKPLANVGVEASY 162
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 112 GNKACGEVYKLGRAKTIADITSAEDKEYGLNNSKYIPTNGMTVGTYFEKIDGLVIGANY 171
OY 163 NYLSEDAANISLG 175
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 172 -LLAQRNKYGTG 183
```

```
RESULT 13
OM25_HAEIN STANDARD: PRT; 371 AA.
AC P46027;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE OUTER MEMBRANE PROTEIN P2 PRECURSOR (OMP P2).
GN OMP2.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
CC Haemophilus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=T1, AND T2;
RX MEDLINE=94293786; PubMed=8022287;
RA Duijn B., van Alphen L., Eljk P., Jansen H.M., Dankert J.;
RT "Antigenic drift of non-encapsulated Haemophilus influenzae major
RT outer membrane protein P2 in patients with chronic bronchitis is
RT caused by point mutations.";
RL Mol. Microbiol. 11:1181-1189(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=T1, AND T2;
RX MEDLINE=94018553; PubMed=8412618;
RA Duijn B., Dankert J., Jansen H.M., van Alphen L.;
RT "Genetic analysis of the diversity in outer membrane protein P2 of
RT non-encapsulated Haemophilus influenzae.";
RL Microb. Pathog. 14:451-462(1993).
CC CC
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
CC -1- SIMILARITY: TO PROTOBACTERIUM STRAIN SS9 OMPH.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC CC
DR EMBL; X73386; CA51803.1; -.
DR EMBL; X73383; CA51800.1; -.
KW Outer membrane; Transmembrane; Porin; Signal.
FT SIGNAL 1 20
FT CHAIN 21 371
FT VARIANT 184 224 K -> Q (IN T2).
FT VARIANT 224 224 A -> T (IN T2).
FT VARIANT 273 273 S -> T (IN T2).
SQ SEQUENCE 371 AA; 40901 MW; D0F252D7A6E5CDA6 CRC64;
```

```
Query Match 11.3%; Score 103; DB 1; Length 371;
Best Local Similarity 24.8%; Pred. No. 0.095;
Matches 53; Conservative 29; Mismatches 74; Indels 58; Gaps 12;
```

```
OY 5 KTLTLLAVSASSLLAMSAANAISYGNADAPYVGAKIGY-----DAKQING-----K 51
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3 KTLAALITVGAFAASANAANAAYVYNNEG-SKVELGGRISLVIAEOSNNTVDOCKQHGALRNQ 61
OY 52 NTAVGIYAGYNDQNF-----GVEAEFVG--SDAKEFNAGVSPVKGVSFGAYGTYRYN 104
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 62 GSRFHFKATHNFGDGYAGYLETRETSIRYQDNADHFD-----DITF-----KYAYV 108
OY 105 FINTPEYAKGKLGIAKTKV-VYTSRNATTYS--NKSQDTSLAGG-VGVGFKPLANVGVA 160
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 109 TLGNKAFGEVKLGRAKTIADITSAEDKEYGLNNSKYIRTNMGNTVGTYFEKID--GLVL 166
OY 161 SYNLSEDA-----NAISLGA 176
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 167 GANYLLAQRDTPANPGKGEVAAGSISNGVOVCA 200
```

```
RESULT 14
ID OM25_BRUAB STANDARD: PRT: 213 AA.
AC 044664;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 25 KDA OUTER-MEMBRANE IMMUNOGENIC PROTEIN PRECURSOR.
GN OMP25.
OS Brucella abortus.
OC Bacteria: Proteobacteria: alpha subdivision: Rhizobiaceae group:
OC Brucellaceae: Brucella.
RN 111
RP SEQUENCE FROM N.A.
RC STRAIN-544S;
RX MEDLINE-95204367; PubMed=7896724;
RA de Weygiffosse P., Lintermans P., Limet J.N., Cloeckaert A.;
RT "Cloning and nucleotide sequence of the gene coding for the major 25-
RL kDa outer membrane protein of Brucella abortus.";
CC J. Bacteriol. 177:1911-1914(1995).
CC -I- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC -I- SIMILARITY: BELONGS TO THE OMP25 / ROPB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X79284; CAA55872.1; -.
CC INTERPRO: IPR000498; -.
CC DR PFAM: PF01389; Ompa-membrane; 1.
CC KW Antigen; Outer membrane; Signal.
CC FT SIGNAL 1 23 POTENTIAL.
CC CHAIN 24 213 25 KDA OUTER-MEMBRANE IMMUNOGENIC
CC FT PROTEIN.
CC SQ SEQUENCE 213 AA; 23052 MW; 23285151F1F794BC7 CRC64;
-----
Query Match 11.2%; Score 102.5; DB 1; Length 213;
Best Local Similarity 27.1%; Pred. No. 0.058;
Matches 60; Conservative 22; Mismatches 60; Indels 79; Gaps 11;
QY 1 MKTLTLAVSSSLIAMSANAISYGNSSADA--QPYVGAKI----- 40
DB 1 MRTLKSLIVSA-ALLPESATAF-----AADAIOQPPVPAPEVAPQYSNAGCYTGILYL 54
QY 41 --GOVDAK-----QINGKNYAGIYAGYNPDON--FGEVEAEFGVSDAKKEFNAGVSPYKG 90
DB 55 GYGMKRAKTSYVSGSIKPDMDKRAGAFAWNNFOODIYVGEVDAGYSNAKKSMDGLEVRKG 114
QY 91 DVKSGAAGCTRYNININPFYAKGKLGIT-----ATKTYDV 125
DB 115 FEGSLGAEVGYDLNLFV-MPYLTAGIAGSQIRLKNGLDESKFRVWGWTAGAGLEAKLTNDNI 173
QY 126 TSR--NATYSNKSDDKSTLAG-----GTVGVGFK 151
DB 174 LGRVEIRYTYGKNK--YDLAGTIVRNKLDIOTDIRVGIYGR 212
-----
RESULT 15
VAC3_HELPY
ID VAC3_HELPY STANDARD: PRT: 1310 AA.
AC 048253;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE VACUOLATING CYTOTOXIN PRECURSOR.
GN VACA.
OS Helicobacter pylori (Campylobacter pylori).
```

```
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
RN 111
RP SEQUENCE FROM N.A.
RC STRAIN-TX30A;
RX MEDLINE-95355366; PubMed=7629077;
RA Atherton J.C., Cao P., Peek R.M. Jr., Tummuru M.K., Blaser M.J.,
RA Cover T.L.;
RT "Mosaicism in vacuolating cytotoxin alleles of Helicobacter pylori.
RT Association of specific vacA types with cytotoxin production and
RT peptic ulceration.";
RL J. Biol. Chem. 270:17771-17777(1995).
CC -I- FUNCTION: INDUCES VACUOLATION OF EUKARYOTIC CELLS. CAUSES
CC ULCERATION AND GASTRIC LESIONS.
CC -I- SUBCELLULAR LOCATION: SECRETED.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U29401; AAA86834.1; -.
CC KW Cytotoxin; Toxin; Signal.
CC FT SIGNAL 1 30 POTENTIAL.
CC CHAIN 31 31 VACUOLATING CYTOTOXIN.
CC FT PROPEP ? 1310 POTENTIAL.
CC SQ SEQUENCE 1310 AA; 141988 MW; 1BC21FE3D435F981 CRC64;
-----
Query Match 11.2%; Score 102.5; DB 1; Length 1310;
Best Local Similarity 22.7%; Pred. No. 0.4;
Matches 48; Conservative 36; Mismatches 74; Indels 53; Gaps 10;
QY 9 AVSASSLLIAMSANAISYGNSSADAQPYVGAKI-----GOVDAKOINCKN 52
DB 1049 AIGGTS-LNNGSNMSTL-YGTSGADVAYLNGEVEAIVGGFGSYGSSFSNQANSINSGANN 1106
QY 53 TAVGIYAGYNPDON-FGEVEAE-FVGSDAKKEFNAGVSPYKGVKSP-----GAVGT 100
DB 1107 TNEGYSRIEFANQHEFDEAGALGSDOSSLNFKSALLQDLNOSTHYLIANSATTTRASTG- 1165
QY 101 YRYNFINTPFYAKGKLGIAKTKVDVTSRNATTYSNKSDDKSTLAGGVGFKPLANVGEA 160
DB 1166 YDFAEFRNALVLKPSVGVSYNHL-----GSTNFKSNSQVALNSGSSQHLFNNANVEA 1220
QY 161 SYN-----ISEDA-----NAISL 174
DB 1221 RYYGDTSYFYMNAGVLOEFARFGSNNNAVSL 1251
```

Search completed: March 6, 2001, 14:40:05
Job time: 173 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 14:38:26 ; Search time 37.09 Seconds
(without alignments)
568.817 Million cell updates/sec

Title: US-09-164-714-7
Perfect score: 912
Sequence: 1 MKTKLLVASASLLAMSA.....SYNYLSEDAANISLGAHLAP 180

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP_TREMBL_15:*
2: SP_archaea:*
3: SP_bacteria:*
4: SP_fungi:*
5: SP_human:*
6: SP_invertebrate:*
7: SP_mammal:*
8: SP_mmc:*
9: SP_organelle:*
10: SP_phase:*
11: SP_plant:*
12: SP_protent:*
13: SP_vertebrate:*
14: SP_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	140.5	15.4	191	2	Q9XD46
2	129	14.1	407	2	P96774
3	122.5	13.4	339	2	Q44298
4	121	13.3	344	2	Q9LA97
5	118	12.9	172	2	Q54470
6	117	12.8	572	2	Q85182
7	117	12.8	572	2	Q30689
8	116	12.7	346	2	Q51841
9	115	12.6	346	2	Q9S539
10	115	12.6	572	2	Q9R952
11	114.5	12.6	160	2	Q9KS75
12	109.5	12.0	250	2	Q46679
13	109.5	12.0	264	2	Q46678
14	109	12.0	165	2	Q33796
15	108.5	11.9	247	2	Q46680
16	107	11.7	174	2	Q9RFV4
17	106	11.6	1291	2	Q9R961
18	105	11.5	356	2	Q9KH07
19	105	11.5	572	2	Q9R949

20	104.5	11.5	357	2	Q48225	Q48225	haemophilus
21	104	11.4	251	2	Q46760	Q46760	escherichia
22	104	11.4	1291	2	Q9ZHU1	Q9ZHU1	helicobacte
23	104	11.4	1291	2	Q9ZHT8	Q9ZHT8	helicobacte
24	104	11.4	1291	2	Q9ZHT5	Q9ZHT5	helicobacte
25	104	11.4	1296	2	Q9ZHT1	Q9ZHT1	helicobacte
26	104	11.4	1296	2	Q9ZHU3	Q9ZHU3	helicobacte
27	103	11.3	403	2	P96773	P96773	haemophilus
28	103	11.3	1291	2	Q9ZHU9	Q9ZHU9	helicobacte
29	103	11.3	1294	2	Q9R959	Q9R959	helicobacte
30	103	11.3	1296	2	Q9ZHT2	Q9ZHT2	helicobacte
31	103	11.3	1296	2	Q9R958	Q9R958	helicobacte
32	102	11.2	378	2	Q9R028	Q9R028	pasteurella
33	102	11.2	1291	2	Q9ZHT9	Q9ZHT9	helicobacte
34	102	11.2	1293	2	Q9ZHT0	Q9ZHT0	helicobacte
35	102	11.2	1296	2	Q9ZHT7	Q9ZHT7	helicobacte
36	102	11.2	1324	2	Q9LBC3	Q9LBC3	helicobacte
37	102	11.2	1324	2	Q9LBC2	Q9LBC2	helicobacte
38	102	11.2	1328	2	Q9LBC8	Q9LBC8	helicobacte
39	101	11.1	572	2	Q9R954	Q9R954	campylobact
40	101	11.1	572	2	Q9R951	Q9R951	campylobact
41	101	11.1	1291	2	Q9ZHT3	Q9ZHT3	helicobacte
42	101	11.1	1291	2	Q9LBC1	Q9LBC1	helicobacte
43	101	11.1	1298	2	Q9ZHT4	Q9ZHT4	helicobacte
44	101	11.1	1303	2	Q9KJ46	Q9KJ46	helicobacte
45	101	11.1	1324	2	Q9LBC7	Q9LBC7	helicobacte

ALIGNMENTS

RESULT 1	
Q9XD46	PRELIMINARY; PRT; 191 AA.
ID Q9XD46	
AC Q9XD46	01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999	(TREMBLrel. 12, last sequence update)
DT 01-OCT-2000	(TREMBLrel. 15, last annotation update)
DE	PURATIVE OUTER MEMBRANE PROTEIN.
OS	Vitreoscilla sp. Cl.
OC	Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
OX	Vitreoscilla.
OX	NCBI_TaxID=96942;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=Cl.
RA	Ma H., Webster D.A., Stark B.C.;
RT	"Region of Vitreoscilla Genome That Rescues Sodium Pumping Deficiency
RT	in E. coli."
RL	Submitted (DEC-1998) to the EMBL/Genbank/DBJ databases.
DR	EMBL; AF14793; AAD41045.1;
DR	HSSP; P36546; 1078.
DR	INTERPRO; IPR000758;
DR	PRINTS; PR00316; ENTEROVIROMP.
DR	PROSITE; PS00695; ENT_VIR_OMP_2; 1.
SQ	SEQUENCE 191 AA; 20939 MW; 04CA84DBF9CACE8F CRC64;
Query Match	15.4%; Score 140.5; DB 2; Length 191;
Best Local Similarity	29.0%; Pred. No. 0.00023;
Matches 58; Conservative 22; Mismatches 91; Indels 29; Gaps 8;	
QY	1 MKTKLLVASASLLAMSAISGNADAPQPYGAKIGGVDAQINGKNTAVIYAG 60
DB	1 MKTKLLVASASLLAMSAISGNADAPQPYGAKIGGVDAQINGKNTAVIYAG 60
QY	1 YN--FDON--FGVEAEFVSGDAKEFNAGVS-----PVKGVKSFQAYGTYRYNFIPTPF 110
DB	1 YN--FDON--FGVEAEFVSGDAKEFNAGVS-----PVKGVKSFQAYGTYRYNFIPTPF 110
QY	55 YNQGQPGWGVNTNTYGGKORNNITSSNOLYENDVDVYSGVGSYR---INPVI 111
DB	55 YNQGQPGWGVNTNTYGGKORNNITSSNOLYENDVDVYSGVGSYR---INPVI 111
QY	111 YAKGKIGIAKTVADVT---SRNATYSNKSDDKTSLAGVGVGKPLANVGEASY----- 162
DB	111 YAKGKIGIAKTVADVT---SRNATYSNKSDDKTSLAGVGVGKPLANVGEASY----- 162
QY	112 NVYGVGVAKADTKGOKONRTGRIVINIDITNVYAGAGVQNPAPNNSVDVGESSHV 171
DB	112 NVYGVGVAKADTKGOKONRTGRIVINIDITNVYAGAGVQNPAPNNSVDVGESSHV 171

QY	163	--NYLSEDDANMISGAHLAF	180
DB	172	NDGYDKRSMNAFNVGXYRF	191
RESULT	2		
P96774			
ID	P96774	PRELIMINARY;	PRT: 407 AA.
AC	P96774;		
DT	01-MAY-1997	(TREMBLrel. 03, Created)	
DT	01-MAY-1997	(TREMBLrel. 03, Last sequence update)	
DT	01-JUN-2000	(TREMBLrel. 14, Last annotation update)	
DE	OMPA2.		
GN	OMPA2.		
OS	Haemophilus ducreyi.		
OC	Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;		
OC	Haemophilus.		
OX	NCBI_Taxid=730;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=35000;		
KX	MEDLINE=91197543; PubMed=9045839;		
RA	Klesney-Tait J., Hiltke T.J., MacIver I., Spinola S.M., Radolf J.D.,		
RT	Hansen E.J.;		
RT	"The major outer membrane protein of Haemophilus ducreyi consists of		
RT	two OmpA homologs."		
RL	J. Bacteriol. 179:1764-1773(1997).		
DR	EMBL; U60646; AAB49274.1; -.		
DR	HSSP; P02934; 1BXW.		
DR	INTERPRO; IPR000498; -.		
DR	INTERPRO; IPR001145; -.		
DR	PFAM; PF00691; OmpA. 1.		
DR	PFAM; PF01388; OmpA_membrane; 1.		
DR	PRINTS; PR01021; OMPADOMAIN.		
DR	PRODOM; PD000930; -; 1.		
QO	SEQUENCE 407 AA; 44702 MW; 0E8E62F3447F8C82 CRC64;		

[illegible]

RN	[1]
RE	SEQUENCE FROM N.A.
RC	STRAIN-NCIMB_1102;
RX	MEDLINE=96198165; PubMed=8626290;
RA	Costello G.M., Vipond R., MacIntyre S.;
RT	'Aetomonas salmonicida possesses two genes encoding homologs of the
RI	major outer membrane protein, Ompa.';
RL	J. Bacteriol. 178:1623-1630(1996).
DR	EMBL: X91983; CAA63036.1; -.
DR	INTERPRO: IPR000498; -.
DR	INTERPRO: IPR001145; -.
DR	Pfam: PF00691; Ompa: 1.
DR	Pfam: PF01389; Ompa_membrane; 1.
DR	PRINTS: PR01021; OMPADOMAIN.
DR	PRODOM: PD000930; -: 1.
KW	Signal.
FT	SIGNAL.
FT	CHAIN
SEQ	SEQUENCE 339 AA: 35807 MW: 71AA46291D3BB7A0 CRC64:

[illegible]

RESULT	4			
09LA97				
ID	Q9LA97	PRELIMINARY;	PRT;	344 AA.
AC	Q9LA97;			
DT	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)			
DT	01-OCT-2000 (TREMBLrel. 15, Last annotation update)			
DE	PUTATIVE OUTER MEMBRANE PROTEIN.			
OS	Aeromonas hydrophila.			
OC	Bacteria; Proteobacteria; gamma subdivision; Aeromonadaceae;			
CC	Aeromonas.			
OX	NCBI_TaxID=644;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-PD134/91.			
RX	MEDLINE-2024644; PubMed-10784058;			
RA	Zhang Y.L., Ong C.T., Leung K.Y.;			
RT	"Molecular analysis of genetic differences between virulent and avirulent strains of <i>Aeromonas hydrophila</i> isolated from diseased fish."			
RL	Microbiology 146:999-1009(2000).			
DR	EMBL; AF146597; AAF45029.2; -.			
SO	SEQUENCE 344 AA; 36096 MW; 82477E05FC07B7A0 CRC64;			

Query Match	13.3%	Score 121;	DB 2;	Length 344;
Best Local Similarity	25.5%;	Pred. No. 0.02;		
Matches 49;	Conservative 30;	Mismatches 59;	Indels 54;	Gaps 9;
QY	1	MKTLETLIAVSASSLLASNAANA--ISYGNSDAQPYYGAK-----IGQYDAKQINGK	51	

Db 2 MKMAPSLIAIAMAAGATAHAADDIYFCAGAGAAHFNGLNKIGGGYAGTEDAAANA - 60
 QY 52 NTAVGYAGYNFQNGVE--AEFVG--SDAKEF-NAGVS-----PVKGDVKSFGAY 98
 Db 61 -----FVQYNTENTENGTFFGYAGRGNTDGLRYENQATISGLAKRLPLGGDFSAFAG 114
 QY 99 GYRYNFIPTPYAKGLIAKTKVDYTSRNATYTSNKSDDKTSLAGVGVGFPLANVCV 158
 Db 115 GAY-----WAHFDGLGTSPTKV-----SPLAGLGVTYKVNALDL 149
 QY 159 EASYNTLSEDAN 170
 Db 150 QARYRYMVDVAD 161

RESULT 5
 Q54470
 ID 054470 PRELIMINARY; PRT; 172 AA.
 AC Q54470;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE OUTER MEMBRANE PROTEIN PRECURSOR.
 GN OMP4.
 OS Serratia marcescens.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Serratia.
 OX NCBI_TaxID=615;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=N28B;
 RX MEDLINE=96036211; PubMed=7582013;
 RA Guasch J.F., Ferrer S., Enfedaque J., Viejo M.B., Regue M.;
 RT "A 17 kDa outer-membrane protein (Omp4) from Serratia marcescens
 confers partial resistance to bacteriocin 28b when expressed in
 Escherichia coli.";
 RT Microbiology 141:2535-2542(1995).
 RL EMBL: Z37157; CAA85513.1; -.
 DR HSSP: P36546; 1008.
 DR INTERPRO: IPR000498; -.
 DR INTERPRO: IPR000758; -.
 DR PFAM: PF01389; Omp4_membrane; 1.
 DR PRINTS: PRO0316; ENTEROVIROMP.
 DR PROSITE: PS00694; ENT_VIR_OMP_1; 1.
 DR PROSITE: PS00695; ENT_VIR_OMP_2; UNKNOWN_1.
 KW Signal; Outer membrane.
 FT SIGNAL 1 24 POTENTIAL.
 FT CHAIN 25 172 OUTER MEMBRANE PROTEIN.
 SQ SEQUENCE 172 AA; 18434 MW; 9E8283AAAE1994DF CRC64;

Query Match 12.9%; Score 118; DB 2; Length 172;
 Best Local Similarity 25.0%; Pred. No. 0.015;
 Matches 48; Conservative 26; Mismatches 86; Indels 32; Gaps 7;

QY 1 MKTLKTLTAVSASSLAMSANAISYNSADAPYVGAIGVDAKQINKNTAVGIYAG 60
 Db 1 MKTIACLSAVAAVY-----AVSAGTAPAGQSTVSAGTAQGDFOGVANKAGFNLYKR 53
 QY 61 YNFDQNFGEVAEFFVGSDAKEFNAGVSPYKGVSEFAGYGYRYNFIPT--PFY----- 111
 Db 54 YEFDNNPLGVIGSF-----THLEKDGSDQDFYKKAQYNSISAGPAVRINDWAS 101
 QY 112 AKKGLGIATKVVYTSRNATYTSNKSDDKTSLAGVGVGFPLANVCVEASY---NTLSED 168
 Db 102 IYGLVDELGKGFNTNANGNSTRHDAD-YGFTYAGAGLQFNPIENVALDVGYEONRIRSYD 160
 QY 169 ANALISGAHLAF 180
 Db 161 VGTWNVGVGYRF 172

RESULT 6

085182
 ID 085182 PRELIMINARY; PRT; 572 AA.
 AC 085182;
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
 DE FLAGELLIN A.
 GN FLAA.
 OS Campylobacter jejuni.
 OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
 OC Campylobacter.
 OX NCBI_TaxID=197;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=D2290;
 RA Meinersmann R.J., Hielt K.L.;
 RT "Concerted evolution of duplicate fla genes in Campylobacter";
 RL Submitted (Feb-1998) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF050188; AAC25641.1; -.
 DR INTERPRO: IPR001029; -.
 DR INTERPRO: IPR001492; -.
 DR PFAM: PF00669; Flagellin_N; 1.
 DR PFAM: PF00700; Flagellin_C; 1.
 DR PRINTS: PR00207; FLAGELLIN.
 DR PRODOM: PD000316; -; 1.
 SQ SEQUENCE 572 AA; 59021 MW; C07AC87B421B59B1 CRC64;

Query Match 12.8%; Score 117; DB 2; Length 572;
 Best Local Similarity 24.5%; Pred. No. 0.081;
 Matches 50; Conservative 24; Mismatches 70; Indels 60; Gaps 7;

QY 22 AATSYNSADAPYVGAIGVDAKQ-----INCKNTAVGIYA----- 59
 Db 242 AAVRAQATSTFAINGVKTKGVKDGANGALVAAINSVYKDTTGVAEASIDANGQLITS 301
 QY 60 ----GYNFDQNFGEVAEFFVGSDAKEFNAGVSPYKGVK-----SFGAYGYRY 103
 Db 302 REGRGIKIDGNITGGA-FTIADMKENVGRSLVKNKGKOLLISGNSLSAGFGA----- 354
 QY 104 NFINTPYAKGLGIKTK--VDYTSRNATYTSNKSDDKTSLAG-----GVGVGFKP 152
 Db 355 ----TFISQASVSLRESKQOIDANIADAMGFGSANKGVYLGYSYSAVMSAGSGCFSS 410
 QY 153 LANVGEASNYNLSEDANAISLGA 176
 Db 411 GSGYSVSGSGKNYSTGFANALTAISA 434

RESULT 7
 Q30689
 ID 030689 PRELIMINARY; PRT; 572 AA.
 AC 030689;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
 DE FLAGELLIN A.
 GN FLAA.
 OS Campylobacter jejuni.
 OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
 OC Campylobacter.
 OX NCBI_TaxID=197;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=D1118;
 RA Meinersmann R.J., Hielt K.L.;
 RT "Concerted evolution of duplicate fla genes in Campylobacter";
 RL Submitted (Feb-1998) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 101-189 FROM N.A.
 RC STRAIN=D224;
 RA Meinersmann R.J., Helsen L.O., Fields P.I., Hielt K.L.;
 RL J. Clin. Microbiol. 0:0-0(1997).

RESULT 10
Q9R952 PRELIMINARY; PRT: 572 AA.
AC Q9R952:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE FLAGELLIN A.
GN FLAA.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter
OX NCBI_TaxID=197;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=117;
RA Meinersmann R.J., Hiett K.L.;
RT "Concatenated evolution of duplicate fla genes in Campylobacter.";
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF050192; AAC25645.1; -
DR INTERPRO: IPR001029; -
DR INTERPRO: IPR001492; -
DR PFAM: PF00669; Flagellin_N; 1.
DR PFAM: PF00700; Flagellin_C; 1.
DR PRINTS: PR00207; FLAGELLIN.
SQ SEQUENCE 572 AA; 59038 MW; F59AA4104EBAZESE CRC64;

Query Match 12.6%; Score 115; DB 2; Length 572;
Best Local Similarity 24.5%; Pred. No. 0.12;
Matches 50; Conservative 24; Mismatches 70; Indels 60; Gaps 7;

QY 22 AATSYGNSADQAPYVAGKIGOVDAKO-----INGKNATYGIYA----- 59
DB 242 AAVRACATSDPRAINCYKIGVDYKDGANGALVAAINSYKDTTGEVASTIDANGOLLIS 301
QY 60 ---GYNFDONFGEAEFVGSAPKFNAGVSPYKGDYK-----SEFAGYGYRY 103
DB 302 RECRGKIDKNTGGCA-FLNADMKENYGRSLYKKGDKILISGSMSSAGFEA----- 354
QY 104 NFNTPFYAKKIGIAKTR--VDVTSRNATTTYSNKSDDKISLAG-----GVGVGFPR 152
DB 355 ---TGFISQASVSLRESGQIDANIDAMGFGSANKGVGLGYSSVAYMSAGSGFSS 410
QY 153 LANVEASVNTSEEDANAISLGA 176
DB 411 GSGYVSGSKNYSSTGRANAIAISA 434
RESULT 11
Q9KS75 PRELIMINARY; PRT: 160 AA.
AC Q9KS75:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE HYPOHETICAL PROTEIN VC1384.
GN VC1384.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Unayam L.A.,
Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
Ermlaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
McDonald L., Uitterback T., Fleischmann R.D., Nierman W.C., White O.,
Salberg S.L., Smith H.O., Colwell R.R., Mekalanos J.D., Venter J.C.,
Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio

RT cholerae.";
RL Nature 406:477-483(2000).
DR EMBL; AE004217; AAF94542.1; -
DR TIGR: VC1384; -
KW Hypothetical protein.
SQ SEQUENCE 160 AA; 16985 MW; 7E47CED047CE409D CRC64;

Query Match 12.6%; Score 114.5; DB 2; Length 160;
Best Local Similarity 29.8%; Pred. No. 0.026;
Matches 56; Conservative 19; Mismatches 70; Indels 43; Gaps 9;

QY 5 KTLAVASASSLLAMSANAAISYGNADQAPYVAGKIGOVDAKONKNTAYGIAGYND 64
DB 3 KTLAL--ALIGASSTA-----MADSMYIGASVGSQSDYECKHG--TAYSVHACTGIL 51
QY 65 QNFGVAEEVYGSAPKFEFA-----GVSPYKDVKSFPACIGYTRNFTNTPYAKGL---- 116
DB 52 PFILGAGYVNHGDFEINATQELASSLYFAVPSMDFGPLH-----VYAKGLHSMD 104
QY 117 -GIATKRVVTSRNATTTYSNKSDDKISLAGVGVGFPLANVGVEASY-NYL--SEDA 172
DB 105 KDINGSKID-----DGLDVMYIGAEVFTIGFPSVGSATYMTYMDSTDVCTL 151
QY 173 SLGAHLAF 180
DB 152 SFNATFHF 159

RESULT 12
Q46679 PRELIMINARY; PRT: 250 AA.
AC Q46679:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE HEAT RESISTANT AGGLUTININ 1.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX Escherichia
RN NCBI_TaxID=562;
RP SEQUENCE FROM N.A.
RC STRAIN=O9:H10:K99;
RX MEDLINE=95012721; PubMed=7927783;
RA Lutyche P., Rupp R., Cavanagh J., Warren R.A., Brooks D.E.;
RT "Cloning, sequencing, and viscometric adhesion analysis of heat-
resistant agglutinin 1, an integral membrane hemagglutinin from
RT Escherichia coli O9:H10:K99.";
RL Infect. Immun. 62:5020-5026(1994).
DR EMBL: U07174; AAC13752.1; -
DR INTERPRO: IPR000498; -
DR PFAM: PF01389; ompA_membrane; 1.
FT CHAIN 26
SQ SEQUENCE 250 AA; 27477 MW; BDDAE89C389BD395 CRC64;

Query Match 12.0%; Score 109.5; DB 2; Length 250;
Best Local Similarity 21.0%; Pred. No. 0.12;
Matches 49; Conservative 38; Mismatches 75; Indels 71; Gaps 10;

QY 1 MKTKTLAVASASSLLAMSANAIS-----VGNADAP 34
DB 1 MEMKNVIAVSALAMGMSSTALADESKTGFVYTKACASVMSLADQRLSCNGEETSK 60
QY 35 YVGAIGOVDAKONKNTAY--GIYAGNFDFONFV---EAEFV--GSDAKEFNAGVS 86
DB 61 YKGG-----DGHDTVSGGIAAGVDEYFQFSIPVTELEFARAKADSKYVWDK 110
QY 87 PVNG-----DVKS-----FGAYGTYRNFNTPFYAKKIGIAK-----TKYDV-- 125
DB 111 SMSGYWRDCLKNEVSVNTLMLNAYYDFRNSAFETWVSAG-IGYAKELHQTGTISTWD 169

OY 126 -----TSRNATYTSNKSDDKTSIAGVGGEFKPLANVGVEASYNVLSSEDAANAIS 173
DB 170 YGCGSSGRESLSRSGSADNFAMSLAGVRYDVTPIALDLSRYLADGSSVS 222

RESULT 13
O46678 PRELIMINARY: PRT: 264 AA.

AC 046678: PRELIMINARY: PRT: 264 AA.

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)

DE HEAT RESISTANT AGGLUTININ 1 PRECURSOR.

OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.

OX NCBI_TaxID=562;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-O9:H10:K99;

RX MEDLINE-95012721; PubMed-7927783;

RA Lutwyche P., Rupp R., Cavanagh J., Warren R.A., Brooks D.E.;

RT "Cloning, sequencing, and viscometric adhesion analysis of heat-resistant agglutinin 1, an integral membrane hemagglutinin from Escherichia coli O9:H10:K99."

RT Infect. Immun. 62:5020-5026(1994).

RL EMBL: 007174; AAC13751.1; -.

DR INTERPRO: IPR000498; -.

DR PFAM: PF01389; OmpA_membrane; 1.

KW Signal.

FT SIGNAL

FT CHAIN

FT SEQUENCE 264 AA: 29024 MW; 335BDAE8E2D4F67 CRC64;

FT SIGNAL

FT CHAIN

FT SEQUENCE 264 AA: 29024 MW; 335BDAE8E2D4F67 CRC64;

Query Match 12.0%; Score 109.5; DB 2; Length 264;
Best Local Similarity 21.0%; Pred. No. 0.13;
Matches 49; Conservative 38; Mismatches 75; Indels 71; Gaps 10;

OY 1 MKTLTLAVSASSLLANSANAIS-----YGSNADAQP 34
DB 15 MIEMKVIIVASALAMAGMSTQALADESKTGFYTGKAGASVMSLADQRFSLGNGEETSK 74
OY 35 YGCAIGQVDAKOINGKNTAY--GIYAGNFDQNGV-----EAEEV--GSDAKKEFNAGVS 86
DB 75 YKGG-----DGHDFVSGIAAGYDFPQESIPVETLEFYARGKADSKYNVDKD 124
OY 87 PVKG-----DVKS-----FGAYGTRYRYNFINTPFYAKKGKLGIAK-----TKVDV-- 125
DB 125 SMSGGYWRDLDLNEVSNTLMLNATYDFRNDSAFTFWVSAG--IGYAKELHQKTTGISIWD 183
OY 126 -----TSRNATYTSNKSDDKTSIAGVGGEFKPLANVGVEASYNVLSSEDAANAIS 173
DB 184 YGCGSSGRESLSRSGSADNFAMSLAGVRYDVTPIALDLSRYLADGSSVS 236

RESULT 14
O33796 PRELIMINARY: PRT: 165 AA.

AC 033796: PRELIMINARY: PRT: 165 AA.

DT 01-JAN-1998 (TREMBLrel. 05, Created)

DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)

DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)

DE ATTACHMENT AND INVASION PROTEIN HOMOLOG.

OS Salmonella typhimurium.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Salmonella.

OX NCBI_TaxID=602;

RN [1]

RP SEQUENCE FROM N.A.

RA De Groote M., Ochsen U.A., Shiloh M., Nathan C., McCoard J.M.,

RA Dinauer M.C., Libby S.J., Vazquez-Torres A., Xu Y., Fang F.C.;

RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF007380; AAB62386.1; -.
DR HSSP: P36546; 1008.
SQ SEQUENCE 165 AA: 17439 MW; 4F998FE466A15B88 CRC64;

Query Match 12.0%; Score 109; DB 2; Length 165;
Best Local Similarity 25.6%; Pred. No. 0.076;
Matches 41; Conservative 19; Mismatches 68; Indels 32; Gaps 6;

OY 36 VGAKIGQVDAKQING-KMT-----AYGIYAGVNF---DQNFGEAEFEV 74
DB 10 VGLALGSLGVRAAGYKNTVSGIAYTDLSGMLSGNANCAKIKYMWEDLDSGFGAMGSVT 69
OY 75 GSDAKKEFNAGVSPVKGVKDFSGAYGTRYN-FINTPFYAKKGKLGIAKRYDVTSRNATY 133
DB 70 YTSADVNNGYKVGADADYTSLLVGSRYFNIDYLN---AYVMIGANAQHI-----KDNM 119
OY 134 SNKSDKTSIAGVGGEFKPLANVGVEASYNVLSSEDAANAIS 173
DB 120 GNSDNKTAFAYGAGIGIQLNPVENIAVNASVEHTSFSTDADS 159

RESULT 15
O46680 PRELIMINARY: PRT: 247 AA.

AC 046680: PRELIMINARY: PRT: 247 AA.

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)

DE HEAT RESISTANT AGGLUTININ 1.

OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.

OX NCBI_TaxID=562;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-O9:H10:K99;

RX MEDLINE-95012721; PubMed-7927783;

RA Lutwyche P., Rupp R., Cavanagh J., Warren R.A., Brooks D.E.;

RT "Cloning, sequencing, and viscometric adhesion analysis of heat-resistant agglutinin 1, an integral membrane hemagglutinin from Escherichia coli O9:H10:K99."

RL Infect. Immun. 62:5020-5026(1994).

RL EMBL: 007174; AAC13753.1; -.

DR INTERPRO: IPR000498; -.

DR PFAM: PF01389; OmpA_membrane; 1.

FT CHAIN

FT SEQUENCE 247 AA: 27104 MW; FFE5D1B261802389 CRC64;

Query Match 11.9%; Score 108.5; DB 2; Length 247;
Best Local Similarity 20.9%; Pred. No. 0.14;
Matches 48; Conservative 38; Mismatches 73; Indels 71; Gaps 10;

OY 4 LKTLAVSASSLLANSANAIS-----YGSNADAQPYVG 37
DB 1 MNKVIIVASALAMAGMSTQALADESKTGFYTGKAGASVMSLADQRFSLGNGEETSKYG 60
OY 38 AKIGQVDAKOINGKNTAY--GIYAGNFDQNGV-----EAEEV--GSDAKKEFNAGVSPVK 89
DB 61 G-----DGHDFVSGIAAGYDFPQESIPVETLEFYARGKADSKRYNVDKDSMS 110
OY 90 G-----DVKS-----FGAYGTRYRYNFINTPFYAKKGKLGIAK-----TKVDV----- 125
DB 111 GGYWRDLDLNEVSNTLMLNATYDFRNDSAFTFWVSAG--IGYAKELHQKTTGISIWDYGY 169
OY 126 -----TSRNATYTSNKSDDKTSIAGVGGEFKPLANVGVEASYNVLSSEDAANAIS 173
DB 170 GSSGRESLSRSGSADNFAMSLAGVRYDVTPIALDLSRYLADGSSVS 219

Search completed: March 6, 2001, 14:38:28

Job time: 131 sec

THIS PAGE BLANK (USPTO)